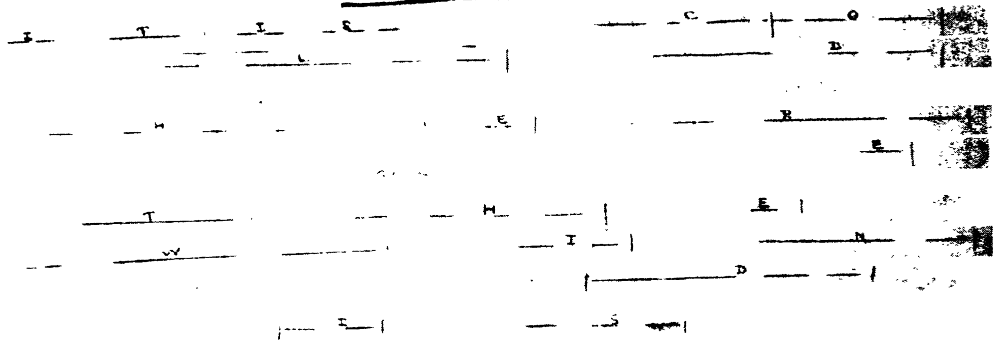
THE ELEMENTS OF
RADIO-COMMUNICATION

Marks made by Marconi's system from the Flatholm Island, at Lavernock Point - May 13th 1897.



Distance about $3\frac{1}{3}$ miles. The first message received by Marconi's [illegible].

[illegible]

[illegible] May. 1897

THE ELEMENTS OF RADIO-COMMUNICATION

BY

O. F. BROWN, M.A., B.Sc. (Oxon.), B.Sc. (Lond.)
LATE SCHOLAR AND RESEARCH STUDENT OF BRASENOSE COLLEGE, OXFORD

AND

E. L. GARDINER, B.Sc. (Lond.)

SECOND EDITION

OXFORD UNIVERSITY PRESS
LONDON NEW YORK TORONTO

OXFORD UNIVERSITY PRESS
AMEN HOUSE, E.C. 4
LONDON EDINBURGH GLASGOW NEW YORK
TORONTO MELBOURNE CAPETOWN BOMBAY
CALCUTTA MADRAS
HUMPHREY MILFORD
PUBLISHER TO THE UNIVERSITY

FIRST EDITION 1927
SECOND EDITION 1939
Second Impression 1943
Third Impression 1945
Fourth Impression 1946

PRINTED IN GREAT BRITAIN

PREFACE TO SECOND EDITION

In the ten years which have elapsed since this book first made its appearance, far-reaching changes have occurred in the art of radio-communication. In particular, the thermionic valve has been developed from one of a number of competing devices used for the detection or production of radio signals, to become the basic device upon which the whole subject largely rests. It is therefore no longer satisfactory to treat the valve late in the book as an alternative to older methods; but it has been thought best to bring forward the chapters in which it is introduced, so that they immediately follow those which deal with basic electrical principles.

In other respects everything possible has been done to retain those features which led to the success of the first edition as a teaching medium. The original clear and simple wording has been retained wherever it remains suitable, as little as possible having been omitted. As a result of this treatment paragraphs may be found in which the wording used is not that in most common use to-day, or in which a subject is dealt with first from the standpoint of early methods, and subsequently from a more modern angle. However, radio-communication can only be understood fully if the historical stages which have led up to its present state are appreciated, and there is no short cut by which these can be neglected. Rather will it be found that they tend to repeat themselves in modernized form. Thus the original ultra-short waves demonstrated by Hertz, after many years in which they were considered useless, are now the latest development used in television broadcasting; whilst the original Fleming two-electrode valve has become the widely used diode detector.

So great is the volume of new material calling for inclusion in this edition, that it has been thought impossible to aim at an advanced technical standard within the space available. The book, therefore, remains an elementary text-book, in which an attempt is made to introduce and explain all aspects of radio-communication as a coherent whole, and without resort to mathematics beyond that required for examination purposes. No attempt has been made to cover receiving equipment in

detail, since countless existing works do this very completely, but due prominence has been devoted to transmission. Particular attention has been paid to a lucid explanation of the principles and reasons underlying radio phenomena, and to their presentation in a style which will appeal to the general reader. For this reason the book is not highly sectionalized, but can be read through more in the manner of ordinary text. If in some cases it is thought that a subject has been treated with insufficient depth, it should be realized that the book is not intended as a detailed work of reference, but as an explanatory introduction which will enable the student to understand and make proper use of more advanced works.

To this end every effort has been made to treat the subject comprehensively, rather than with great depth. The valve chapter has been much extended, and contains a description of certain special types, e.g., the cathode-ray tube and electron multiplier, not usually found in elementary works. The reader's attention has been directed to future developments whenever possible, chapters having been added on Television both from its present aspect and from that of the several unsuccessful attempts which have paved the way to the excellent results now obtained. A chapter has been added explaining the important subject of aerial design, while reference will be found to recent special developments such as aerial navigation by radio, ultra-short wave working, and the quartz-crystal filter.

In response to the suggestions of those using the first edition as a teaching medium, a selection of examination questions have been added at the end of each chapter, and thanks are due to those institutions which have kindly given permission for the reprinting of these from their examination papers. The book is intended to cover the syllabus of the City & Guilds examination in radio-communication, Grade 1 (preliminary grade). It should also prove helpful to students entering for the Graduate of the Institute of Electrical Engineers examination, and for the examinations set by the Institute of Wireless Technology, both in radio-communication and in television. The latter examinations, however, are of a more advanced nature and are more closely related to current commercial practice than most, and it is not, therefore, claimed that the book would provide sufficient information for these unless augmented by an extensive study of current periodicals and publications. This is particularly the

case with television, for which it is doubtful whether a comprehensive text-book yet exists. A number of questions from these more advanced examinations have been included after some chapters, and whereas they are in all cases dealt with in the text, there will be a few which cannot be completely answered by students without additional reading or assistance from the tutor.

E. L. G.

CONTENTS

	PREFACE	v
I.	HISTORICAL AND INTRODUCTORY	1
II.	PROPERTIES OF HIGH-FREQUENCY ALTERNATING CURRENTS	14
III.	PRODUCTION OF ELECTRICAL OSCILLATIONS AND RADIATION OF ELECTRICAL ENERGY	37
IV.	TRANSMISSION OF DAMPED WAVES AND THE RESONANCE OF TUNED CIRCUITS	56
V.	THERMIONIC VALVES AND THE CATHODE-RAY TUBE	70
VI.	TRANSMISSION OF CONTINUOUS WAVES	117
VII.	DETECTION AND RECTIFICATION	165
VIII.	AMPLIFICATION OF ELECTRICAL OSCILLATIONS	198
IX.	MODULATION AND RADIO-TELEPHONY	227
X.	RECEIVING CIRCUITS	258
XI.	SELECTIVITY IN RADIO RECEPTION	320
XII.	FAITHFUL REPRODUCTION IN RADIO-TELEPHONY	350
XIII.	PROPAGATION OF WAVES THROUGH SPACE	393
XIV.	DIRECTIONAL RECEPTION AND THE AERIAL SYSTEM	424
XV.	BASIC PRINCIPLES OF TELEVISION	481
XVI.	PRINCIPLES OF MODERN TELEVISION WORKING	523
	INDEX	547

CHAPTER I

HISTORICAL AND INTRODUCTORY

WIRELESS, mainly through its application to broadcasting, probably enters more intimately into everyday life than any of the achievements of physical science during the last thirty years. Over seven million people in Great Britain listen nightly in their own homes to the programmes transmitted by the British Broadcasting Corporation. Speech transmitted from an ordinary telephone on the desk of an office in New York is heard as clearly in London as if the speaker had been talking from Birmingham or Manchester. Messages are transmitted daily by radio-telegraphy from Europe to the farthest corners of the earth. Yet the pioneers upon whose labour this great structure of commercial enterprise is built never dreamed of any one of these applications. Their whole interest was in science for its own sake, and their work, such as the laboratory experiments of Hertz and the theoretical and mathematical achievements of Clerk Maxwell, provides a convincing answer to those who question the use of pure scientific research. Without the mathematical predictions of Clerk Maxwell, Hertz would never have carried out his experiments or realized what their results indicated. Without the work of Maxwell and Hertz, Marconi would not have set out to apply their laboratory results to the problem of practical communication.

Previously to Maxwell's time the theories put forward to explain the radiation of heat and light provided no satisfactory explanation of the phenomena observed. In fact, the theories advanced were so self-contradictory that they were worse than useless. Maxwell showed that light and heat were radiated through space by vibrations which constitute electromagnetic waves: or in other words, that a body giving out light sets up rapidly-changing electric and magnetic forces in the medium surrounding it, which in turn cause similar forces to be produced, so that vibrations spread out all round the body until they reach our eye and make themselves perceptible as light. In much the same way, if a stick is moved backwards and forwards in a pond, the motion of the stick is communicated to the water and is conveyed through it, so that corks, for

example, some way off are set in vibration by the ripples or waves resulting from the motion of the stick.

Maxwell further demonstrated that the only difference in the propagation by radiation of heat and light lay in the fact that the vibrations we call light had a greater frequency or shorter wavelength than those which make themselves felt as heat. He predicted, however, that still slower vibrations than those perceived as heat must exist. These slower vibrations were detected by Hertz in the later eighties and were applied by Marconi to radio-telegraphic communication.

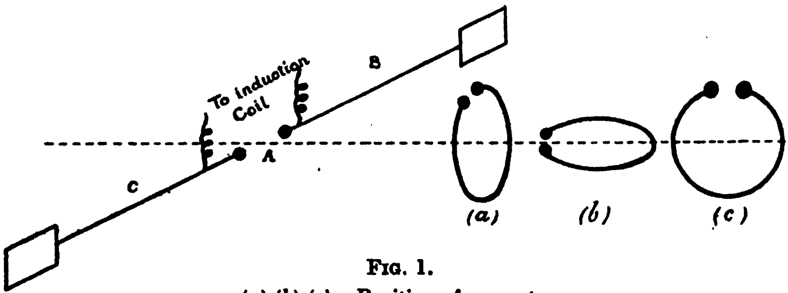


FIG. 1.
(a) (b) (c)—*Position of resonator.*

Hertz's experiments were briefly as follows: He attached two equal rods ending in metal plates to the terminals of the secondary of an induction coil, as shown in Fig. 1. He connected a battery to the primary terminals (not shown in the figure) and caused sparks to take place at A between metal balls at the ends of the rods B and C. This arrangement he called an oscillator. He then took what he called a resonator, which consisted of a wire bent into a circle ending in two small metal balls whose distance apart could be adjusted. This he held in different positions along the line through A at right angles to the rods. He found, when the diameter of the resonator and the size of the spark gap in it were carefully adjusted, that in certain positions sparks were produced between the balls of the resonator. Thus in position (*a*) Fig. 1, sparks were obtained in the position shown, but no sparks were obtained when the resonator was turned round in its own plane so that the spark gap of the resonator was at right angles to that of the oscillator; in position (*b*) sparks were obtained at the resonator when its spark gap faced the oscillator, but as the resonator gap was turned away from the oscillator

the sparks decreased in strength and ceased when the gap was farthest removed from the oscillator gap; in position (*c*) no sparks were observed whatever the position of the resonator gap. These results he explained on the hypothesis that the electric forces radiated from the oscillator built up electric forces in the resonator, when set in a favourable position, which were capable of producing sparks across the balls. He proved by further experiments that he was actually detecting the electromagnetic waves predicted by Clerk Maxwell, for he found that the vibrations sent out from his oscillator travelled with the same velocity as light. In addition, he showed that they could be reflected and refracted in the same way as light waves, for while the electromagnetic waves could pass through a non-conductor and could be refracted, for example, by a prism of paraffin wax, they were reflected by conductors such as large sheets of metal or metallic gauze.

With the Hertzian resonator the electric waves could be detected at very short distances only. During the early nineties, however, many scientists, including Branly and Lodge, experimented with a more sensitive type of detector, to which the general name *coherer* was given. In its usual form this type of instrument consisted of a glass tube containing two pieces of metal between which were a large number of small metallic particles. Wires connected to the two pieces of metal were brought out through the glass and the tube was exhausted and sealed. If such a tube is connected with a battery and a galvanometer, only a very small deflection is observed on the galvanometer scale, since the resistance of the metal particles to the flow of the current is very great. If, however, electric oscillations are passed through the coherer, the particles adhere together, the resistance drops, and the galvanometer is deflected. When the coherer is shaken, the resistance again increases. Thus, to use the instrument for the continuous detection of groups of oscillations, an arrangement was provided by which it was automatically kept sensitive by being repeatedly tapped by a hammer like that of an electric bell. This was termed a decoherer.

It was with an instrument of this nature that Marconi attempted to receive signals sent by electric waves over distances of several miles. His first great advance resulted from his earthing one end of the Hertzian oscillator and extending

the other end vertically into the air as an aerial. His receiving apparatus consisted of a vertical oscillator, similar to his transmitting aerial, in which the coherer was substituted for the spark gap. An audible method of detecting the signals by means of a telephone was used instead of the galvanometer. It was soon realized that it was better not to introduce the spark gap and the detector directly into the sending and receiving circuits, but to excite the aerial by inducing oscillations through the mutual action of a coil in the aerial and another coil placed in a closed circuit containing the spark gap and a condenser; and likewise to place the detector in a secondary circuit similarly coupled to the receiving aerial.

The next major discovery was due to Sir Oliver Lodge who realized the importance of tuning or synchronizing the circuits employed, i.e. of arranging that the circuits should all resonate to the particular frequency of the oscillations which it was desired to send out. He did this by adjustment of the size of the coils and condensers employed. As a result of the improved sensitiveness thus attained Marconi, in 1901, was able to detect signals in Newfoundland sent out from Cornwall. His apparatus was roughly of the type described, and the power input 20 kw., which for many years after this first transatlantic communication remained a low power for that distance. Until recent years powers of 1,000 kw. and over have been necessary for reliable long-distance communication, and it is only with the development of short-wave directional beam signalling that the use of lower power has become possible.

In the Marconi apparatus used in these experiments the condenser, introduced as described into the spark-gap circuit, was charged by current from an induction coil or a transformer until a spark occurred across the spark gap. While the spark is taking place, the resistance of the spark gap is comparatively low, and the condenser therefore discharges across it and causes high-frequency oscillations to be set up in the aerial, which gradually die away.

The number of these oscillations taking place per second depends upon the adjustment of the aerial and condenser circuits. The manner in which these oscillations are set up by the discharge of the condenser will be considered later; but it may be remarked here that such oscillations produce what are known as damped waves. A group of these oscillations is

produced whenever the induction coil or transformer charges the condenser to a sufficient degree to break down the resistance of the air between the spark balls. This may take place many times in a second. A dot or a dash of the Morse code transmitted by means of damped waves will, therefore, consist of a series of groups of high-frequency oscillations following each other, as shown in Fig. 2. The oscillations S_1, S_2, etc., composing each group G_1, G_2, etc., may be taking place at the rate of hundreds of thousands per second, while the groups G_1, G_2, G_3, etc., may follow each other either regularly or irregularly at a rate of perhaps 100 to 1,000 per second.

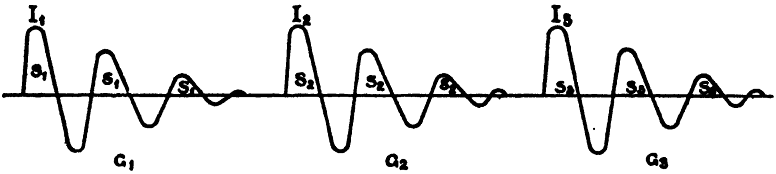


FIG. 2.

In the twelve years from 1900 the line of development of radio-telegraphy was mainly towards the improvement of spark systems for the emission of damped waves. The objects striven for consisted in: (1) increasing the number of high-frequency oscillations contained in each group by decreasing the rate at which oscillations died away; (2) making the number of groups of high-frequency oscillations follow each other regularly, so that a given number were sent out per second.

We shall see later that actually each group G_1, G_2, G_3 produces a sound in the receiving telephones, so that, if the number of groups emitted per second is regular, a musical note will be heard which, for various reasons, is better for reception than the unmusical sound produced when the groups do not follow each other regularly.

In England spark transmission was improved by the introduction of the Marconi rotating spark gap, which both decreased the rate of decay or damping of the oscillations and gave a musical note in the telephone receiver. In Germany the Telefunken Company developed the quenched spark system, in which the same objects were achieved by means of a specially designed spark gap. These methods of transmission are still

infrequently used by ships at sea for emergency purposes, but will be finally abandoned in 1938.

On the receiving side the method of detecting the waves was improved by the introduction of crystal detectors and by the Fleming two-electrode thermionic valve. This latter was the forerunner of the three-electrode valve which was later to revolutionize both transmission and reception.

By 1912 the Marconi Company was conducting a commercial service between Clifden and Glace Bay in Canada, and at the beginning of the war a large 300-kw. spark station had been completed near Carnarvon for communication with the United States of America. The development of radio-communication between ships at sea and between ships and the shore was also far advanced; and the number of lives saved by wireless at sea was already very large.

It was realized early in the development of radio-communication that many advantages could be obtained from the use of continuous or 'undamped' waves, i.e. waves the amplitude of which remained constant, so that a dot or dash would consist of a series of steady oscillations, instead of a series of groups of oscillations, each of which died down practically to zero before the next group started. Compared, however, to the progress made with transmitting systems employing damped waves, little progress had been made by 1914 in the development of systems employing continuous waves. This was largely due to the fact that no entirely satisfactory method of receiving such waves was then available.

Poulsen and Pedersen in Denmark had, however, developed to a considerable extent the Poulsen Arc system of transmission, in which oscillations are produced, as described in a later chapter, by connecting the aerial to one side of an electric arc, the other side being connected to earth. The arc system was further developed in America by C. F. Elwell, and during and since the War many of his high-power arcs were installed in stations in Europe, including Rome and Lyons. Two of the most recent arc stations were the British Post Office Stations at Leafield and Cairo.

In addition to the arc system, one or two high-frequency alternating machines, for the direct production of oscillations of radio-frequencies, had been designed prior to the War. Later the design of such machines was carried to a very efficient

stage of development, especially by Latour in France and by Alexanderson in America, and machines were installed in many French and American high-power stations. This method of transmission was not, however, taken up to any large extent in this country, and is now entirely superseded by valve transmission.

During the War the most remarkable advance was the improvement of the three-electrode thermionic valve as an instrument both for the reception and for the generation of oscillations. The year or two preceding the War had seen the introduction of the Meissner valve in Germany and the Round valve in England, and also of the 'audion' invented by Dr. Lee de Forest in America. These early valves contained a small trace of gas, such as hydrogen or helium, or mercury vapour. They were irregular in their action and required constant skilled adjustment. During the War, however, the French produced a 'hard' valve in which the exhaustion of the glass containing-vessel was carried to the highest possible extent. These valves were much more steady and regular in their action and, although many advances have since been made in the design of valves, yet the majority of those made to-day are, in principle, very similar to the French 'R' valves as regards the theory of their action.

It was early found that the hard valve could be applied to the aerial circuit for the amplification of weak signals which would be totally inaudible with the detecting apparatus previously used. It was also found that several valves could be used in cascade, i.e. the signal as amplified by the first valve could be applied to the next, and the increased amplification thus obtained communicated to a third, and so on. In this way the oscillations picked up by the aerial could be amplified several hundred times before being applied to the detector. If, even then, the amplified and detected signals were still too weak for convenient reception, additional valves could be introduced for the magnification of the low-frequency pulses produced by the detector before these were applied to the telephones. Thus, amplifiers containing perhaps seven valves in cascade had been evolved, in which the first three acted as high-frequency amplifying valves, the next as a detector, and the remaining three as low-frequency magnifiers. For various reasons, which will be dealt with in later chapters, it was usually impossible to employ

more than seven valves arranged as indicated above. But as each of the amplifying valves, both for high and low frequency, augments the strength of the received signal, it is clear that such a receiver may be perhaps a thousand times more sensitive than a single-valve or crystal receiver.

With amplifiers of this nature it was found possible to alter considerably the type of aerial employed for reception. Instead of high and long receiving aerials, it was possible to make use of aerials wound round rectangular frames whose sides measured only a few feet. A frame aerial possesses very marked properties of directive reception. When its plane is placed parallel to the direction of propagation of the waves which are being received, oscillations of maximum strength are excited, but when its plane is at right angles to the direction of propagation of the waves, nothing is heard in the receiver.

Further, by the use of a frame aerial suitably orientated, it was found possible to arrange for a receiving station to be erected close to a transmitting station, in such a way that reception could be carried out while the near-by station was working at full power. Thus a simple and satisfactory system of duplex working was made possible.

Again, by the use of a frame aerial which could be turned about a vertical axis, it was found possible to determine the apparent direction of a given transmitting station. This fact is the basis of the various systems of directional reception now employed as aids to navigation both by air and sea.

Thermionic valves also provided a very satisfactory means of generating continuous waves of all frequencies, and valve transmitters of various powers, from a few watts to many hundreds of kilowatts, were designed. The waves generated by valve transmitters are very pure and easily controlled; and, for this reason, the introduction of the valve gave an immense impetus to radio-telephony with its later application to broadcasting. In America the trend of invention has been towards the development of single valves, made partly of glass and partly of metal, each capable of generating very powerful oscillations. In this country, on the other hand, general practice has tended towards the employment of several valves in parallel, the component valves used in powerful transmitters being usually made of silica and capable of dealing with perhaps 10 to 15 kw. each. The American method in which very large

single valve units are employed is gaining ground in our own stations of recent years however, together with the steady improvement in output and cost of large water-cooled valves, which are now constructed largely of metal and can be taken apart for service purposes.

A revolution in radio-communication technique dates from about 1924, when Marconi conducted experiments from the yacht *Elettra* to determine the possibilities of short-wave channels replacing the existing long-wave systems. It was found that with a decrease in wavelength below those generally employed at the time an improvement in the daylight range occurred, and this unexpected improvement led to the first daylight communication with Sydney on a wavelength of only 32 metres. Previously to this the best contacts had been from Poldhu to South America on a wavelength of the order of 100 metres, when it had been found possible to reduce the power to as little as one kilowatt, but at the expense of pronounced fading. The success of this work led the Marconi Company to accept a contract for the erection of a series of beam stations for services to the Dominions, and to guarantee to the Post Office a working reliability previously considered impossible. In so doing Marconi placed his faith in the improvement which would result from the use of effective directional or 'beam' aerial systems, able to concentrate the waves from comparatively low-powered transmitters and direct them towards the destination for which they were intended. The work of Hertz had shown that electromagnetic waves could be concentrated by reflectors of suitable design in the same way as light waves; but such reflectors are only of manageable dimensions when the wavelengths are short, say less than 200 metres at most. Whilst to Marconi, therefore, must go the credit of first realizing the practical possibilities of the beam systems foreshadowed by the theoretical work of Hertz, it is probable that his assistant Franklin has done more than any other scientist to put directional radiation upon a practicable basis, and to work out the design of commercially effective aerial arrays by which the necessary concentration can be utilized.

Late in 1926 the first beam circuit from England to Canada was opened, and proved a striking success. It was followed soon afterwards by others to Australia, South Africa, and India, by the inauguration of the short-wave broadcasting stations at

Chelmsford and Pittsburg, and radio-telephone services to ships at sea. From this period onwards progress in the development of short-wave communication became rapid and the lead given by the Marconi Company was followed in America and other countries. The reliability and economy of short-wave working became so apparent that the development of long-wave circuits largely ceased, and considerable competition began to be felt for the first time from radio by the many established cable companies. In this country the position was eased through the formation of the vast merger termed Imperial and International Communications Ltd., in which the leading radio and cable interests are now merged under a degree of government control.

Mention must now be made of the extremely successful experiments in radio-communication carried out by British and American amateurs, and which went a long way to show commercial engineers that possibilities were being overlooked in the short-wave region. After the War amateur stations which had previously been allowed longer wavelengths were subjected to a degree of control, and forced to rely solely on wavelengths below 200 metres, then considered useless for serious purposes. Following on the most commendable persistence and endeavour, amateur signals were first received in England from America in 1921 on wavelengths slightly below 200 metres, and by 1924 communication in both directions on shorter wavelengths and lower powers had become frequent. In the same winter a British amateur, Mr. Goyder, exchanged the first short-wave signals with New Zealand; and thus it is seen that short-wave communication was both discovered and to some extent developed by those working essentially for the advancement of science.

In the years between 1927 and the present time progress has been on so wide a front that it is difficult to select any individual steps for consideration above others. The trend will become evident in more detail as the following chapters are read, whilst the daily and periodical press has done much to make modern radio achievement familiar to every reader. Basic theory has changed little in its essentials since the adoption of short waves, but there has been a considerable advance in details of design, and in efficiency. Probably the single factor having greatest influence on the technique of short-wave working has been the steady evolution of the valve, from the early diode and triode

to complex types of ten or more electrodes, and from power handling capacities of a few watts up to single valves handling 500 kw. or more. Circuit design has naturally tended to follow the improved variety and capabilities of valves, and in some cases has been delayed until the most suitable types have become available; whilst in other cases new circuit conceptions have evoked special forms of valve. Valve types frequently employed in a common circuit have been constructed as a combined unit, contained in a single evacuated envelope, thus reducing cost and simplifying equipment. Certain new derivatives from the radio valve such as the gas-filled triode and the cathode-ray tube have made their appearance, and have evolved to a position of far-reaching importance. At the present time the phenomena of secondary emission from photo-electrically sensitive surfaces, previously considered a defect in valve construction, has led to the Zworikin electron multiplier; a device of such enormous amplification and low noise level that it may lead in the next few years to a further drastic revision of radio-communication methods, comparable to that brought about by the original introduction of the valve, or of short-wave working.

Notwithstanding the practical developments of recent years, however, it is becoming increasingly clear that many problems of radio-communication can only be solved by an extension of our fundamental knowledge of the physical phenomena underlying them.

For instance, the problem of mitigation of the disturbances produced by electric discharges in the atmosphere, known as 'atmospherics', still remains largely unsolved in spite of the repeated claims of inventors. These disturbances form a great handicap to radio-communication, and hundreds of methods have been suggested for overcoming them. Some of these have proved totally useless; others have had a limited success; but none has proved satisfactory when applied to the elimination of the very strong atmospherics met with in the tropics. This failure may be attributed mainly to the fact that, until recently, little has been known of the fundamental nature or the origin of the disturbances it was desired to eliminate. The Radio Research Board of the Department of Scientific and Industrial Research in this country has been the first to take up this problem with any real success, and the extremely important

results obtained by Watson Watt and Appleton will be described later.

The variations in the strength of signals is another case in point. Signals from a transmitting station, when received at a distance, vary in strength within wide limits, from season to season, from day to day, or even after dark from minute to minute, quite apart from all changes in the transmitter or receiver. Moreover, with waves of medium wavelength, stations which may be totally inaudible by day give strong readable signals by night. Without some explanation of the cause of these variations, or at least the collection of reliable data concerning them, the scientific design of transmitting stations would be impossible. The variations are connected with the electrical state of the upper atmosphere and with the effect of the surface of the earth on wave propagation. Experiments indicate that these factors affect long waves very differently from short waves, and their investigation has reached the stage when the relative behaviour of different wavelengths can be foretold with reasonable but not perfect accuracy. It is now becoming possible to estimate in advance the most suitable wavelength which should be employed for communication between any two places at any time of the year and at any part of the twenty-four hours. Such computations are becoming sufficiently accurate to form the basis of design in commercial radio-communication, and hence it can be said that the vagaries of the upper atmosphere are now moderately well understood.

These variations in strength are popularly classed under the heading of fading and were for a long time a source of serious detraction from the enjoyment of long distance broadcast listening. Their effect has been considerably reduced, however, by the recent introduction into receiver design of 'automatic volume control', by which the sensitivity of the receiver is automatically increased to compensate for any reduction in the amplitude of the incoming waves. Unfortunately, fading is often accompanied by distortion, introduced into radio-telephone signals through the unequal fading of the various components which compose them. This effect is reduced by diversity reception, in which the signals received from a number of independent aerials situated several wavelengths apart are combined to provide the final programme. Since fading will not in general be simultaneous at each of the several aerials, their combined

effect is comparatively uniform, and it is by the use of this method that the B.B.C. can effect the relay of foreign programmes at good quality. No equally effective palliative yet exists, however, in the case of the domestic receiver.

With regard to direction-finding, it has been found that sudden and considerable variations occur at night in the observed apparent direction of the received waves. These variations, which in certain circumstances may render direction-finding unreliable, are due to much the same factors that cause variations in the strength of signals. Thus, here again, further improvements must be looked for in the application of knowledge gained from the fundamental study of the physical phenomena of propagation. Just as in the beginning pure physics made radio-communication possible, so after more than a quarter of a century of organized effort it is through pure physics that the next great advance must be sought.

CHAPTER II

PROPERTIES OF HIGH-FREQUENCY ALTERNATING CURRENTS

IN order to understand how communication can be effected by radio-telegraphy it is necessary to consider what happens in the transmitting apparatus, how the energy created is radiated into space, how it is propagated through space, and how it finally influences the receiving apparatus. Both the transmitting and receiving apparatus consist of more or less complicated electric circuits, and as a first step it is necessary to consider the laws which govern the flow of electricity in these circuits and, in particular, those which apply to what are known as high-frequency currents.

Suppose we have a body charged with electricity and we bring a body charged with like electricity up to it from a distance; then, as the two bodies will repel one another, a certain amount of work must be done. The amount of work necessary to bring a small body charged with unit quantity of electricity up to the first body from an infinite distance is defined as the *potential* of the first body. In this case an infinite distance is taken to be a distance so great that the initial effect of the charge of the first body on our unit charge is so small as to be negligible.

Now when water flows through a pipe, the amount of water which will flow in a given time depends on the hydraulic pressure at the ends of the pipe. So in a similar way the current of electricity which will flow through a conducting wire depends on the difference of electrical pressure at the ends of the wire. This electrical pressure is represented by the difference of potential between the ends of the wire.

If we make the necessary potential difference at the ends of a wire by connecting them to an accumulator or a dry battery, we obtain what is known as a *direct current*, i.e. a current which flows always in the same direction and is constant in strength from moment to moment as long as the potential of our cell remains steady and the circuit is left unchanged. In radio-telegraphy, however, we have to deal not with currents and potentials which remain constant from moment to moment, but with high-frequency oscillating currents and potentials, i.e.

currents and potentials which, starting from zero, increase to a maximum, decrease again, reverse and increase to a maximum in the other direction, and decrease to zero, the whole process being repeated perhaps tens or hundreds of thousands of times per second. Similar alternating currents are employed for many industrial purposes, but in such cases the number of complete cycles through which the current goes is usually only fifty per second.

To understand the effects of the high-frequency currents of wireless telegraphy, it is necessary to grasp the ideas expressed by the words self-induction, mutual induction, and capacity.

SELF- AND MUTUAL INDUCTION

Suppose we consider the simplest possible electrical circuit, that shown in Fig. 3, in which the battery B is joined to the two ends of a resistance R through a key K. After the key is closed, suppose we have a difference of potential E, due to the voltage or electromotive force (E.M.F.) of the battery, applied to the ends R, which will drive a current I through the circuit. The quantities E, I, and R are connected by a relation known as Ohm's Law, viz. $E = IR$. In this relation R is expressed in ohms, E in volts, and I in amperes.

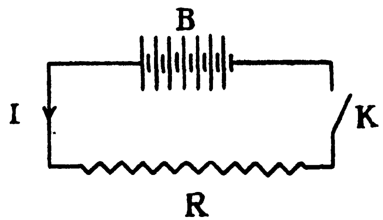


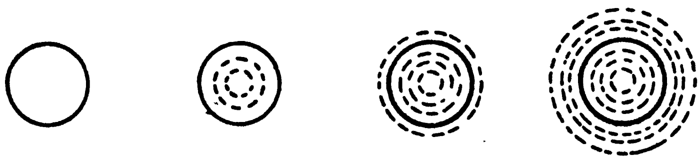
FIG. 3.

The work done in t seconds in driving the current of I amperes through the resistance R ohms is EIt. This work will appear as heat in the wire.

When a current flows along a wire, a magnetic field, which we can represent by lines of magnetic force, is created round the wire. This magnetic field arises in the centre of the wire and spreads outwards. When the key K (Fig. 3) is closed, the current through the wire will not reach a steady value immediately, but will grow gradually, and the magnetic field will similarly grow. If we represent the lines of magnetic force by circles, we get at different intervals of time concentric rings of magnetic force spreading outwards, as represented by the dotted lines in Fig. 4. Here the full line represents the cross-section

of the conductor and the dotted lines represent the growth of the magnetic flux at various instants until the current reaches its steady state, as shown in Fig. 4 (*d*).

If the magnetic field round a conductor is changing, or if lines of magnetic force are moving across it, a potential difference will be induced across the conductor. In the present case, the wire is being cut by a varying magnetic field moving out from the centre. The effect of this changing field is to produce an E.M.F. in opposition to the E.M.F. of the battery, and at



(*a*) (*b*) (*c*) (*d*)

Fig. 4.

any instant the current through the wire is that due to the difference of these two E.M.F.s. When the current reaches its steady value the strength of the magnetic field becomes constant also, and the opposing or back E.M.F. vanishes.

When the key is opened and the current begins to die away, the magnetic lines of force collapse again towards the centre of the wire, and an E.M.F. is again induced in the wire which will tend to maintain the current at the value it had before the key was opened. The direction of the E.M.F. induced by the changing of the magnetic field will always be in a direction such as to oppose any change in the current.

The effect we have been considering is known as self-induction. The amount of self-induction in a circuit depends on the way the conducting wires of the circuit are arranged. In the case of a straight wire, the lines of force from one piece of the wire cannot link up or cut any other piece, and the effect of self-induction is small. If the wire is bent back on itself, the effect of self-induction is eliminated altogether. On the other hand, if the wire is bent into a coil or helix, so that the magnetic lines from any one turn interlink with the other turns, its effect is very great. If an iron core is introduced into the coil so that the magnetic flux is further increased, the effect of self-induction is greater still. Since the flux in the iron core

is not directly proportional to the current, the effective self-induction varies with the current.

The practical unit of self-induction is called the *henry*; and a coil of wire is said to have a coefficient of self-induction of one henry when, if the current through it changes at the rate of one ampere per second, an E.M.F. of one volt is induced in the coil.

The calculation of the self-induction of a coil of any pattern is usually a difficult matter, because it is necessary to take into consideration the effects of the magnetic lines of force due to any one turn upon all the other turns. In many practical books tables and curves for the calculation of different types of coils are given.

Hitherto we have considered only one example of the effect of self-induction, namely, its effect at the making and breaking of a continuous current. With the high-frequency oscillating currents and potentials of radio-telegraphy, however, the magnetic fields due to the currents are continuously varying very rapidly, so that the effect of self-induction becomes very important.

From the definition already quoted it can be seen that the induced E.M.F. across an inductance of L henries is equal to $-L \times$ rate of change of the current in amperes per second.

The negative sign is present because the E.M.F. of self-induction is opposed to the applied E.M.F. whose direction is regarded as positive. From this relation it can be shown that the maximum value V of the induced E.M.F. in the case of alternating currents is given by

$$V = \omega LI.$$

In this formula I is the maximum value of the alternating current and $\omega = 2\pi f$, where f is the frequency of the current, i.e. the number of complete cycles or oscillations taking place in a second.

It will be seen that the above relation has a form similar to Ohm's Law, except that, instead of the resistance R in ohms, we have a quantity ωL. This quantity is known as the reactance of the coil and, like R, is expressed in ohms. If the frequency of the oscillations is very large, ωL may be much greater than the ordinary resistance R of the wire of the coil, which may then often be neglected compared to the reactance of the coil.

When a magnetic field is produced round a wire a certain amount of energy is necessary to create it. When the magnetic field collapses, as the current decreases, this energy is restored to the circuit again.

Suppose we have a current which increases steadily from zero to a maximum of I amperes in t seconds; then its rate of change is I/t, and if it flows through a coil of self-induction L the back E.M.F. of self-induction will be $L \times I/t$.

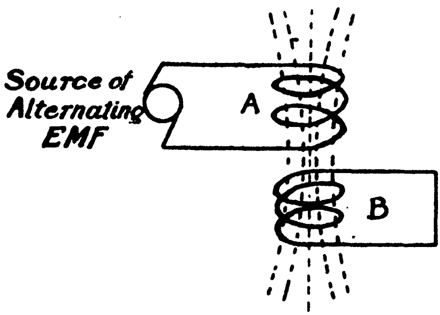


FIG. 5.

Now the work done in t seconds, as we saw on page 15, is given by the product of the E.M.F., the current, and the time. In the present case the current is not steady. If, however, we suppose that it varies uniformly from 0 to I, we can take its average value as $\frac{1}{2}I$. Hence we get the value of the work done in forcing this quantity of electricity against the back E.M.F., or, what is the same thing, the quantity of energy stored in the magnetic field, as:

$$L\frac{I}{t} \times \frac{I}{2} \times t$$

$$= \tfrac{1}{2}LI^2.$$

Next, let us suppose we have a circuit in which the current is changing, and that we have a coil A of self-induction L in this circuit, and that the lines of magnetic force, due to the varying current, thread or link not only the coil itself A, but a second coil B in another separate circuit close by (see Fig. 5). Then, since the magnetic field through B is varying, an E.M.F. must be induced in the second circuit. This E.M.F. will create a current in B which in turn produces a magnetic field. As in the case of self-induction, the direction of the E.M.F. will be

such as to produce a magnetic field which will tend to keep the current in A from altering its value.

Thus, the inductive effects in two circuits react upon each other. This interaction is termed *mutual induction.* It is clear that the amount of interaction between the coils will depend on the number of lines of magnetic force due to A which can link B; and, analogously with what we said in the case of self-induction (which is indeed a special case of mutual induction), we may say that a pair of coils have a coefficient of mutual induction of one *henry* when a current in one coil, changing at

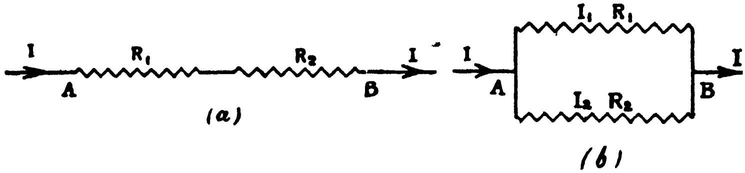


FIG. 6.

the rate of one ampere per second, induces an E.M.F. of one volt in the other coil. The mutual induction of a pair of coils is usually denoted by the letter M. Two circuits are said to be *coupled* together magnetically by means of linkage of the magnetic lines of force due to the currents in the coils, i.e. by their mutual induction. The coils are said to be loosely or tightly coupled according as to whether their mutual inductance is small or great.

If we have two resistances in series in a circuit, such as R_1 and R_2 in Fig. 6 (a), and if I is the current and E the potential difference across them between the points A and B, then by Ohm's Law

$$E = IR_1 + IR_2$$

or

$$E = I(R_1 + R_2),$$

and therefore the total resistance of the two resistances R_1 and R_2 in series is equal to $R_1 + R_2$.

If the resistances are connected in parallel as in Fig. 6 (b), the current I will be divided between them. Let I_1 be the current flowing in the resistance R_1 and I_2 that in R_2. Then if the fall of potential between the ends of the resistances A and B is E, we have

$$E = I_1 R_1 \quad \text{and} \quad E = I_2 R_2;$$

also

$$I = I_1 + I_2;$$

and therefore $$\frac{E}{R_1}+\frac{E}{R_2} = I_1+I_2 = I.$$

$$\therefore\ E\left(\frac{1}{R_1}+\frac{1}{R_2}\right) = I.$$

Hence if the equivalent resistance of the two resistances is R, this equivalent resistance is given by

$$\frac{1}{R} = \frac{1}{R_1}+\frac{1}{R_2}.$$

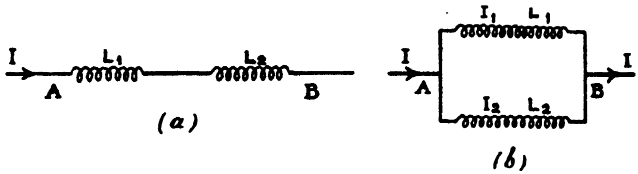


FIG. 7.

Similarly, suppose we have two inductances L_1 and L_2 connected in series, as in Fig. 7 (a), and an alternating potential, whose maximum value is V, and whose frequency is $f = \omega/2\pi$, is applied between A and B. Let I be the maximum value of the high-frequency current produced by the alternating potential. Then since, as we have seen, the reactance of the two coils to currents of a frequency corresponding to ω is ωL_1 and ωL_2, we have the relation

$$V = \omega L_1 I+\omega L_2 I,$$

and hence the equivalent self-induction L of the two coils in series is given by $$L = L_1+L_2.$$

As before, if the two coils are in parallel as in Fig. 7 (b), then the current divides into two portions I_1 and I_2 in the two coils, such that $$I = I_1+I_2.$$

Also $$V = \omega L_1 I_1 \text{ and } V = \omega L_2 I_2,$$

$$\therefore\ \frac{V}{\omega L_1}+\frac{V}{\omega L_2} = I_1+I_2 = I,$$

$$\therefore\ V\left[\frac{1}{L_1}+\frac{1}{L_2}\right] = \omega I,$$

and therefore the equivalent self-induction of the two coils in parallel is given by

$$\frac{1}{L}=\frac{1}{L_1}+\frac{1}{L_2}.$$

If we consider again the effects of the mutual induction between the coils A and B in Fig. 5, it is clear that the mutual induction of the coils A and B will cause the currents produced in coil B to react on coil A. The effect of this reaction will depend on the direction of the lines of magnetic force due to B which link A. The direction of the lines depends on the direction of the current in B, i.e. on the direction in which the spiral of wire composing B is wound. If B is wound in a direction such that the magnetic force due to B is in the same direction as that due to the current in A, then the mutual induction of B will help any self-induction effect which A may have. If, however, the direction of the winding of B is reversed, the mutual induction due to B will oppose the self-induction of A.

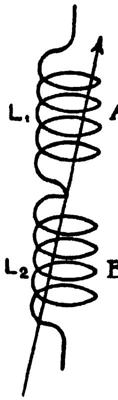


FIG. 8.

These facts are employed in a type of variable inductance coil sometimes used in wireless circuits and known as a *variometer*. In this instrument two coils are arranged so that one can rotate inside the other. The coils are also connected in series. In one position the magnetic fields of the two coils add together, but if the moving coil is rotated through 180° the magnetic fields will oppose each other. In the first position the mutual induction of the two coils will in the case of each coil be *added* to the self-induction of the coils: in the second position it must be *subtracted*. Let the coils be A and B, and let the arrangement be represented graphically in Fig. 8, where B is supposed to be the moving coil. Let the coefficients of self-induction of the coils be L_1 and L_2 and let the mutual induction between the coils be M. Then since the coils are connected in series, the equivalent reactance of the whole instrument, when the directions of the windings of A and B are similar, will be

$$(L_1+M+L_2+M)\omega$$

or

$$(L_1+L_2+2M)\omega.$$

When, however, the coil B is rotated through 180°, the equivalent reactance of the instrument will be

$$(L_1-M+L_2-M)\omega$$

or

$$(L_1+L_2-2M)\omega.$$

When the coil B is at right angles to A, no lines of force due to one coil can cut the other, and therefore M is zero, and the inductance of the pair is L_1+L_2.

The variometer thus provides a variable inductance which can have any value of equivalent self-induction between the values L_1+L_2+2M and L_1+L_2-2M, depending on the position into which B is rotated.

Capacity

If an insulated conductor is joined by a conducting wire to another conductor which is charged with electricity, then a current of electricity will flow until the potential of the two conductors becomes the same.

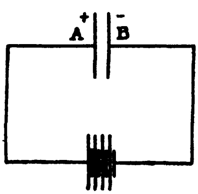


Fig. 9.

Any conductor has capacity, i.e. the power of collecting a charge or storing up electrical energy, depending on its size and shape. When the charge on a body is increased, more work is necessary in bringing another unit of like charge up to it from a great distance—in other words, its potential has been raised. The quantity of electricity necessary to raise the potential of a body by unity is defined as the capacity of the body. If two metal plates, one charged positively and one charged negatively with electricity, are brought near to one another as in Fig. 9, then, through the influence of the charges on the two plates, their potential difference is reduced, and to bring it back to its original value a quantity of electricity must be added to the plates.

This arrangement is obviously one whereby an increased capacity can be obtained, and is called a *condenser*. At the present time there is a movement to substitute the word 'capacitor' for condenser, and 'capacitance' for capacity, thus bringing them into line with inductance or resistance. These terms may be met with on occasions, but it is doubful if they are becoming generally popular. Condensers may be of many

types, as, for example, parallel metallic cylinders or concentric spheres, but usually they consist of a number of parallel metal plates separated by air or some other insulating material such as ebonite or mica—alternate plates being joined together by conductors. The insulating material between the plates is called the 'dielectric'. The greater the area of the plates and the nearer they are together the greater is the capacity of the condenser. The expression for the capacity of a parallel plate condenser is C (farads) $= KA/4\pi d$, where A is the total area between the plates in sq. cm.; d the distance between them in cm.; and K the Permittivity or Specific Inductive Capacity of the dielectric.

Suppose the two sets A and B of alternate plates of such a condenser are joined to the poles of a battery of voltage E, the plates A to the positive pole and the plates B to the negative pole, a momentary current will flow so that the plates A become positively charged and the plates B negatively charged. The potentials of the plates A and B will finally become the same as of the poles of the battery. The unlike charges on A and B will attract each other across the dielectric separating them, which will therefore be in a state of strain. In other words, energy will be stored in it. If we disconnect the battery and connect the plates A and B by a wire, a current will again flow until the charges are neutralized and the energy stored up is dissipated. The action is somewhat analogous to that of a compressed spring. The more the spring is compressed the more mechanical energy is stored up, which is released when the spring is allowed to uncoil.

The capacity of a condenser is usually denoted by the symbol C. The unit of capacity is the farad, and a condenser is said to have a capacity of one farad when one coulomb* of electricity changes its potential by one volt.

If we suppose that Q coulombs of electricity will raise the potential difference across a condenser of capacity C farads through E volts, we get the general formula $Q = CE$.

We can deduce the amount of energy stored in the dielectric of a condenser of capacity C farads, when it is charged to E volts, as follows: Suppose the condenser is charged in time t; then, since $Q = CE$, the average current flowing into the

* A coulomb is the unit of quantity of electricity, and is the quantity of electricity carried by one ampere flowing for one second.

condenser, i.e. the quantity of electricity flowing per second, is CE/t; also, if the potential changes steadily from zero to E, we may assume its average value to be $E/2$. Therefore the work done in charging the condenser in time t, or the energy stored when the charging is complete, is

$$\frac{CE}{t}\times\frac{E}{2}\times t = \tfrac{1}{2}CE^2.$$

If several condensers are joined in parallel as in Fig. 10 (a),

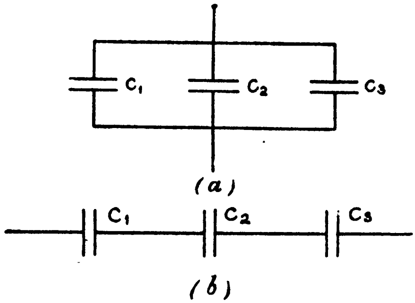


FIG. 10.

the voltage across them will be the same, and if Q_1, Q_2, Q_3 is the charge on each and C_1, C_2, C_3 their capacities, then

$$Q_1 = C_1E, \qquad Q_2 = C_2E, \qquad Q_3 = C_3E.$$

$\therefore$ If Q is the total charge present,

$$\begin{aligned} Q &= Q_1+Q_2+Q_3 \\ &= C_1E+C_2E+C_3E \\ &= (C_1+C_2+C_3)E. \end{aligned}$$

Thus the joint capacity of the three condensers in parallel is equal to the sum of the capacities of the individual condensers.

If the condensers are joined in series, as in Fig. 10 (b), the charge Q on the condensers will be the same, so that

$$E_1 = \frac{Q}{C_1}, \qquad E_2 = \frac{Q}{C_2}, \qquad E_3 = \frac{Q}{C_3},$$

and the total potential drop across the three condensers is

$$E = E_1+E_2+E_3;$$

$$\therefore\ E = Q\left[\frac{1}{C_1}+\frac{1}{C_2}+\frac{1}{C_3}\right].$$

Thus the combined capacity of the three condensers in series is given by

$$\frac{1}{C}=\frac{1}{C_1}+\frac{1}{C_2}+\frac{1}{C_3}.$$

It will be seen that the combined capacity of two condensers in series is less than that of the individual condensers. Comparatively little change is made in the capacity of a small condenser if a large condenser is placed in series with it.

Graphic Representation of Alternating Currents

From what has already been said it will be seen that an alternating current may be represented by the curve on the

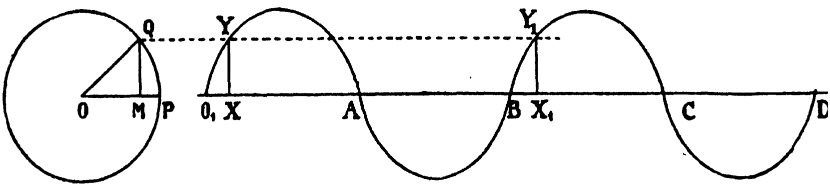


FIG. 11.

right-hand side of Fig. 11. The strength of the current at various instants of time is shown by the ordinates of the curve, while intervals of time are shown along the horizontal axis. The current increases to a maximum, and then decreases to zero at A, where it changes sign. It then increases to a maximum in the other direction, and completes a whole cycle in a time represented by O_1B. From B to D the cycle is repeated again. Thus the current XY at time O_1X is exactly the same and is in the same direction as the current X_1Y_1 at time O_1X_1. As previously stated, the frequency, which we will call f, is the number of cycles completed in one second, or the number of times the current passes through points such as B and D in one second. The period of the current T is the time taken by the current in passing through one cycle and is obviously equal to $1/f$.

At the left-hand side of the figure another way of representing the variations of the current is shown. Here OP may be supposed to be a rotating rod of length equal to the maximum value of the current. If OP rotates with uniform angular velocity, so that it turns through an angle of 360° (or, what is the same thing, 2π radians) in time T (the period of the alternating current), we see that OP will make a complete revolution while the

current goes through a complete cycle. In one second OP will turn through an angle of $2\pi f$ radians, where f is the frequency of the current. The current at any moment is given by the perpendicular drawn to OP from the point on the circumference of the circle which the rod has reached. After OP has turned through an angle θ, for which a time $T \times \theta/2\pi$ will be taken, the current is represented by MQ which is equal to XY or $X_1 Y_1$; after OP has turned through 90°, or $\pi/2$ radians, the current is a maximum and then decreases until it passes through zero when OP has turned through 180° or π radians. A wave of the shape shown in Fig. 11 is known as a 'sine' wave. The wave form of an oscillation may be very different from a 'sine' wave, but it can be shown mathematically that a wave of any form, however complicated, can be regarded as made up of a number of 'sine' waves of different amplitudes and frequencies, suitably related to each other in phase. The process of analysing a complex wave into its component sine waves is termed a Fourier Analysis.

Phase Difference

There is one other term used in connexion with alternating currents which must now be explained, and that is *phase* or *phase difference*. Suppose on the same diagram (see Fig. 12) we represent two currents, or a current and an E.M.F., having the same frequency but different maximum values. Suppose also that they do not pass through their zero values at the same instant, but that the current curve indicated by the continuous line $OABC$ lags behind the dotted curve $O_1A_1B_1C_1$. Then when the current represented by $OABC$ is passing through the point O, the current represented by $O_1A_1B_1C_1$ has a value OY. Similarly when the current $O_1A_1B_1C_1$ is passing through zero at B_1, the value of the current $OABC$ is given by B_1Y_1 flowing in the opposite direction to OY. Thus the current represented by $OABC$ lags behind that represented by $O_1A_1B_1C_1$ by an interval of time equal to O_1O.

If we represent the alternating current or E.M.F. $OABC$ by a circle described by a rod O_2P, and $O_1A_1B_1C_1$ by a circle described by a rod O_2P_1, O_2P lags behind O_2P_1 by an angle θ. This angle represents the phase difference between the first current and the second current. If this angle were 90° or $\pi/2$,

then the two quantities would be 90° out of phase and the value of one would be zero when the other was a maximum, and vice versa. If the phase difference were 180° or π, the two quantities would be exactly opposed to one another. If θ is zero or a multiple of 360°, then the currents are said to be in phase. It can be seen that θ is given by $\frac{O_1O}{O_1A_1} = \frac{\theta}{\pi}$.

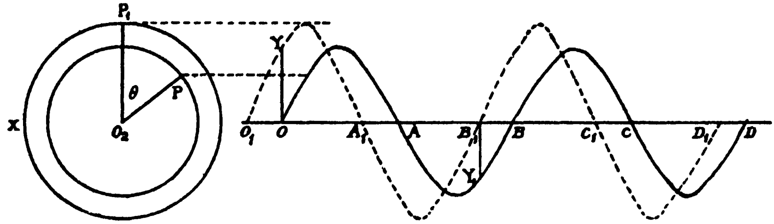


FIG. 12.

APPLICATION OF HIGH-FREQUENCY E.M.F.S TO CIRCUITS

Let us now consider what happens when an alternating high-frequency potential of maximum value V is applied by some means to a coil of self-induction L. An alternating current of the same frequency as the E.M.F. will flow through the coil. Let its maximum value be I. Both V and I are changing from moment to moment, but for our purposes it will be sufficient to consider their maximum values only. Now we have seen that the effect of a rapidly-changing current is to create a changing magnetic field, which in its turn will induce an opposing E.M.F. due to self-induction in the coil. From the curves shown in Fig. 13 (*a*) it will be seen that the *rate of change* of the current, and therefore the rate of change of the magnetic field created (see Fig. 13 (*b*)), is greatest when the current is passing through its zero value. The induced E.M.F., which is equal to $-L\times$ rate of change of current, must therefore have a maximum value when the current is zero, and have a zero value when the current is a maximum. This opposing or back E.M.F. is therefore 90° out of phase with the current, and, as already stated, has a maximum value of ωLI. The back E.M.F. can therefore be represented by the full line in Fig. 13 (*c*). In mechanics it is well known that action and reaction are equal and opposite.

If, for example, we try to push a railway truck along a line, the reaction of the truck will be equal at every moment to the force we apply to it. Hence in the same way the applied voltage

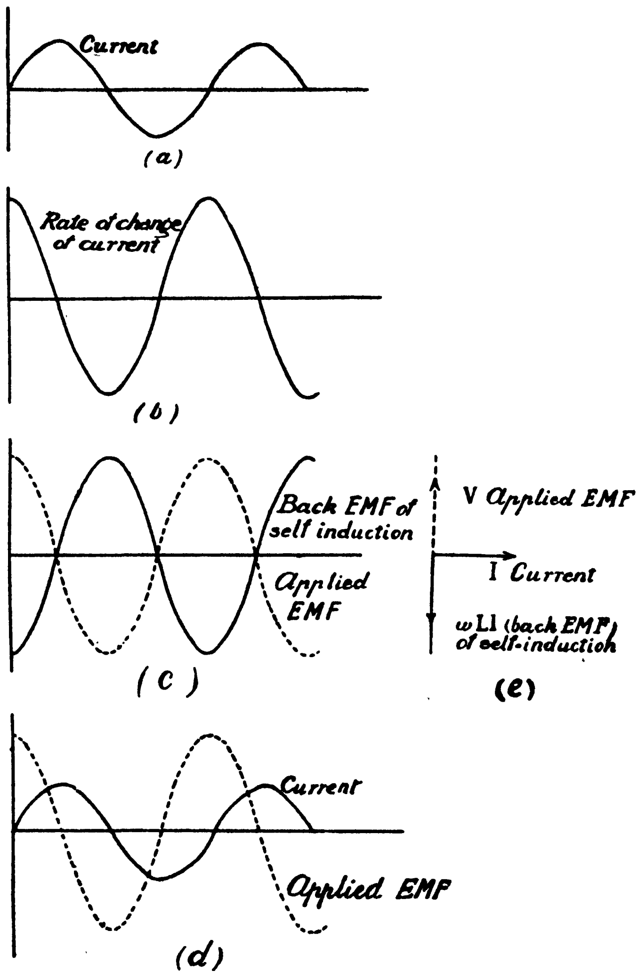


FIG. 13.

must balance at every moment the back E.M.F. due to the self-induction of the coil. The applied voltage will thus be represented by the dotted line of Fig. 13 (*c*). The current and the applied E.M.F. must therefore be 90° out of phase, as shown in Fig. 13 (*d*). Its maximum value *V* must be just sufficient to

equal the product of the reactance of ωL ohms multiplied by the maximum value of the current I.

We may represent the applied E.M.F. V and the induced E.M.F. ωLI by two equal straight lines, as shown in Fig. 13 (*e*), which would rotate with equal angular velocities to form circle diagrams of alternating currents. A diagram such as that in Fig. 13 (*e*) is known as a vector diagram.

Circuit containing Coil and Resistance. Hitherto we have supposed there was no resistance in the circuit; but let us now

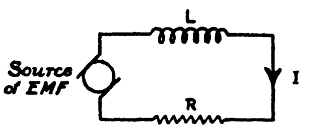


FIG. 14 (*a*).

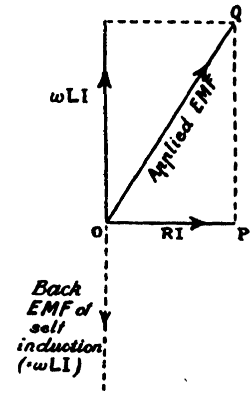


FIG. 14 (*b*).

consider the case in which an alternating E.M.F. V is applied in a circuit containing both a coil of self-induction L and a resistance R, as shown in Fig. 14 (*a*). Once again the applied E.M.F. must provide a component ωLI to overcome the reactance due to the coil. This will be 90° out of phase with the current. It must also provide a component *in phase* with the current to drive the current through the resistance R. This component by Ohm's Law will be equal to RI. Thus the applied E.M.F. must provide two components which are 90° out of phase with each other (see Fig. 14 (*b*)). These cannot be added algebraically, but must be added geometrically, just as we add forces in the parallelogram or triangle of forces in mechanics. The applied E.M.F. will therefore be given by the hypotenuse OQ of the right-angled triangle OPQ, or

$$OQ^2 = QP^2+OP^2$$
$$V^2 = \omega^2L^2I^2+R^2I^2$$
$$V = I\sqrt{(\omega^2L^2+R^2)}.$$

We therefore see that the impedance (i.e. the equivalent resistance) of the circuit is equal to $\sqrt{(\omega^2L^2+R^2)}$ ohms.

Impedance, Conductance, and Admittance. The term Impedance, to which we have just been introduced, is an important one. It is the factor in an alternating current circuit which corresponds to the resistance of a direct current circuit, and takes into account both the inductance, capacity, and ohmic resistance. It is thus a complex quantity, involving the ideas of phase, which we have seen implies vectorial methods of addition or subtraction. The term Reactance is used when we wish to describe the equivalent resistance to alternating currents of a pure capacity or a pure inductance. A reactance, therefore, contains no ohmic resistance. Should resistance be appreciable, Impedance would be the correct term to use. On the other hand, the impedance of a practical circuit will be composed of two parts, the ohmic resistance which has no effect upon phase and can be dealt with arithmetically; and the capacitive or inductive reactances which must be treated vectorially.

It is sometimes convenient to express the conductivity of a circuit as a definite quantity. It is then given the name Conductance, is defined as the reciprocal of resistance, and measured in 'reciprocal ohms', or 'mhos'. Similarly, the reciprocal of Inductance is given the name Admittance, and is also expressed in mhos. When we are dealing with a complex Impedance or Admittance, it may be convenient to treat the Conductance of the resistive elements separately. In this case the reciprocal of the capacitive and inductive impedances is termed the Susceptance of the circuit. These several expressions should not be regarded as of prime importance. They are names assigned for convenience to different aspects of the same physical property, that of resistance or impedance, and are not widely used in elementary work.

Condenser in a High-frequency Circuit. We have seen that when a battery is connected to a condenser a current will momentarily flow into the condenser plates, until they receive a charge such as to make the difference of potential between them equal to that of the voltage of the battery. The current will then cease. If, however, an alternating high-frequency source of E.M.F. is connected to the condenser, the plates will collect a charge until the applied E.M.F. reaches a maximum and begins to decrease, when the condenser will begin to dis-

charge. When the alternating E.M.F. passes through the zero value, the condenser will be completely discharged and will then begin to collect a charge of opposite sign to that of the former charge, since the E.M.F. has reversed its direction. If the condenser has first stored positive electricity on one set of plates, it will now collect negative electricity on this set. This cycle of operations will continue for each cycle of the alternations of the E.M.F.

We have been considering above the charge or the quantity of electricity on the plates. Now the current, as we have seen, is the rate of change of the quantity of electricity, and, since the quantity of electricity is varying from moment to moment, we see that, when the condenser is attached to a source of alternating potential, an alternating current of electricity will flow through the circuit. Thus a condenser does not prevent the flow of an alternating current as it prevents the flow of a direct current.

When the plates are collecting a charge, the presence of this charge alters the potentials of the plates of the condenser in such a way as to produce an E.M.F. in opposition to the applied E.M.F. The maximum value of this opposing E.M.F. can be shown to be equal to $I/(C\omega)$, C being the capacity of the condenser, I the maximum value of the current, and $\omega = 2\pi f$, where f is the frequency of the applied E.M.F. Just as the self-induction L of a coil produces a reactance of $L\omega$, so a condenser introduces in the circuit a reactance of $1/(\omega C)$ ohms, and, as action and reaction are equal and opposite, the applied E.M.F. must balance the opposing E.M.F. due to the charge on the plates of the condenser.

The rate at which the quantity of electricity on the plates of the condenser is changing will be greatest when it is beginning to increase from zero or is passing through its zero value. Hence the current will be greatest when the applied E.M.F. is zero, and will itself be zero when the applied E.M.F. is a maximum. Thus the difference in phase between the E.M.F. and the current will be 90°. The current will, however, be leading the E.M.F., instead of lagging behind as in the case of a coil. Hence, if in a vector diagram we represent the current by a horizontal line OC drawn towards the right as in Fig. 15, the applied E.M.F. will be represented by a line OP drawn downwards equal to $I/(\omega C)$ and at right angles to OC. The opposing E.M.F. of the

condenser will be represented by $OP_1 = OP$ drawn vertically upwards. The opposing E.M.F. of the condenser is thus exactly opposite in phase to the opposing E.M.F. due to the self-induction of a coil.

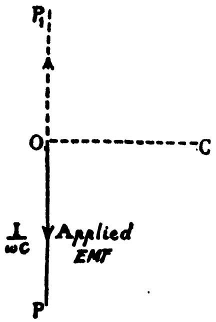


FIG. 15.

Circuit containing Condenser, Coil, and Resistance. Hence, if we consider an alternating E.M.F. applied to a circuit containing a condenser, a coil, and a resistance, as shown in Fig. 16 (a), we see that, in order to produce a current of maximum value I in the circuit, the applied E.M.F. must provide:

(1) A component which we represent by a horizontal line equal to RI to drive the current through the resistance R;

(2) A component ωLI to overcome the reactance ωL due to the coil (this E.M.F. we can represent by a line drawn vertically upwards);

(3) A component $I/(\omega C)$ represented by a line drawn vertically downwards to overcome the reactance of $I/(\omega C)$ of the condenser.

Since these last two components are in the same straight line, they can be added algebraically.

The applied E.M.F. will be given by the hypotenuse OQ of the right-angled triangle OQP, as shown in Fig. 16 (b), in which PQ ($= OQ_1$) represents $\{L\omega - 1/\omega C)\}\ I$, and OP represents IR.

If the maximum value of the applied E.M.F. is V, then

$$V = I\sqrt{\left\{R^2 + \left(L\omega - \frac{1}{C\omega}\right)^2\right\}},$$

or the impedance of the circuit is equal to

$$\sqrt{\left\{R^2 + \left(L\omega - \frac{1}{C\omega}\right)^2\right\}} \text{ ohms.}$$

Circuits of this nature, consisting of a coil and a condenser, are among the most important met with in many arrangements of radio transmitting and receiving apparatus.

The alternating source need not be introduced into the circuit, but may be applied by induction from a neighbouring circuit,

or, as in some types of transmitting apparatus, by a discharge across a spark gap.

If we wish to obtain the greatest possible oscillating current from a given applied E.M.F. of frequency given by ω, we must arrange that the impedance of the circuit is as small as possible. If the resistance of the circuit remains the same, we can obtain the smallest impedance by arranging the self-induction of the

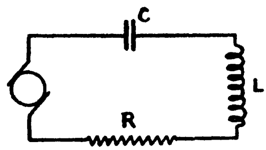


FIG. 16 (*a*).

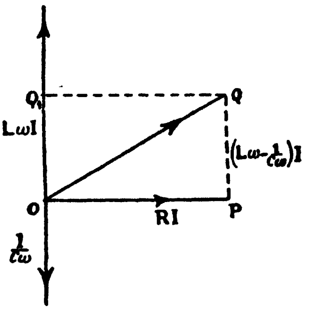


FIG. 16 (*b*).

coil and the capacity of the condenser in such a manner that the reactance is zero, i.e.

$$L\omega - \frac{1}{C\omega} = 0,$$

or

$$\omega^2 LC = 1,$$

or, since $\omega = 2\pi f$,

$$f = \frac{1}{2\pi\sqrt{(LC)}}.$$

In this case the component of the applied E.M.F. necessary to overcome the reactance of the coil is just equal to the component necessary to overcome the reactance of the condenser. Therefore the only work the applied E.M.F. has to do is to drive the current through the ohmic resistance of the circuit, and thus we are able to obtain as large a current as possible with the given applied E.M.F.

The condition described above is known as *electrical resonance*, and the frequency f given by $f = 1/2\pi\sqrt{(LC)}$ is known as the *resonance frequency*.

This expansion also represents approximately the *natural frequency* of the circuit, which, as will be shown later, is the

frequency at which the circuit itself will tend to oscillate, or to which it will tune when used in a receiver.

Power Factor. When a direct current flows through a resistance across which there is a potential difference, the power is given by the product of voltage and current, being $I \times V$ watts. This is also the case in an alternating current circuit containing a pure resistance only, for the current and potential will be in phase, and can therefore be multiplied arithmetically. Hence, when the load circuit contains resistance only, the alternating current power in it will be the product of the alternating current and voltage.

When a circuit contains inductance or capacity, however, the case is different. For example, an alternating current flowing through either a pure inductance or a pure capacity containing no resistance will be exactly 90° out of phase with the alternating potential causing it to flow. The vector product of the two is then zero, and no power whatever is dissipated in the circuit. This fact may seem strange at first, but it should be borne in mind that a circuit without resistance seldom occurs in practice, and some power will therefore be lost. It is, however, possible for a current of hundreds of amperes to flow through a large condenser of good quality without more than a few watts of power being dissipated. Such a current is sometimes termed a 'wattless current'.

In practice the power developed in a condenser or inductance depends upon the phase angle, and is given by $VI \cos \phi$, where ϕ is the angle by which the current leads or lags behind the voltage. This factor $\cos \phi$ is termed the 'Power Factor' for this reason, that it defines the proportion of the possible power that is actually developed in any reactance. In a complex impedance the power developed in the resistive and reactive elements can be calculated separately and added, because power itself involves no idea of phase. We are concerned with alternating power in radio-transmission, for example, when considering a radiating aerial. The aerial is a complex impedance, but it is usually possible to replace it by an 'equivalent resistance' as explained later. When measuring aerial power it is usual to replace the aerial by an actual equivalent resistance, in which the alternating power is dissipated as heat. The latter can then be measured experimentally, or the current and voltage which are now strictly in phase can be read by suitable instruments.

EXAMINATION QUESTIONS

1. A circuit consisting of a resistance of 50 ohms, an inductance of 10 henries, and a capacitance of 1 microfarad in series is found to pass 1 ampere with a certain applied E.M.F. alternating at 50 cycles per second. What will it pass with the same E.M.F. alternating at 100 cycles per second?
City and Guilds Institute. Preliminary Exam. 1935.

2. A variable condenser has a maximum capacity of 1,000 micromicrofarads and a minimum capacity of 100 micromicrofarads. When in the maximum position it is charged to 1,000 volts. The charging supply is then disconnected and the condenser turned to the minimum position. What was the original energy and the final energy stored in the condenser? Explain the reason for any difference. *C. and G. of L. I. Preliminary Exam. 1936.*

3. An inductance coil has an inductance of 0·025 henry. When a 100-volt 50-cycle potential is impressed across the coil, the current flowing is 10 amperes. What are the resistance and power factor of the coil? *C. and G. of L. I. Preliminary Exam. 1937.*

4. What is meant by phase difference and power factor of a condenser? A current of 10 amperes (R.M.S.) at a frequency of 500 kilocycles is flowing through a 0·01 mfd. condenser. Calculate the power dissipated in the condenser, if its power factor is 0·02.
Institute of Wireless Technology. May 1936.

5. Derive from first principles an expression for the combined capacity of a number of condensers joined in series. Will the proportional change in the combined capacity be greatest if an additional condenser be added in series, when the latter is of large or small capacity?

6. What is meant by the dielectric constant of a material? In what way does it vary? A condenser has a capacity of 1,000 mmfd. and a resistance of 2 ohms. Calculate its power factor at a frequency of 100 kilocycles. *A. I. W. T. June 1937.*

7. A condenser consisting of eleven rectangular plates, each two inches by one inch separated by mica plates 4 mils thick, has a capacitance of 4,000 mmfd. What is the dielectric constant of the mica? *Grad. I. E. E. 1935.*

8. A resistance of 1,500 ohms and an inductance of 5 henries are connected in parallel across a 50-cycle-per-second alternating current supply of 1,000 volts R.M.S. What will be the total current taken from the mains? *C. and G. Grade 1. 1933.*

9. What is meant by the permittivity of a dielectric? An air dielectric fixed condenser and inductance are found to resonate at a wavelength of 500 metres. The condenser and inductance are immersed in oil, and it is then found that they resonate at 750 metres. What is the permittivity of the oil?

C. and G. Preliminary. 1935.

10. What is a variometer, and for what purposes has it been used in radio-circuits? State the relationships, giving the combined inductance of two coils connected (*a*) in series, and (*b*) in parallel, when there is no magnetic coupling between them. How does the presence of such a coupling modify the combined inductance?

11. What were the important steps in the history of radio-communication which led up to the use of resonant circuits at both transmitter and receiver?

12. Explain the terms 'Conductance' and 'Admittance'. A circuit comprises a resistance of 100 ohms in series with a condenser of 1 mfd. and an inductance of 10 henries. What will the conductance of this circuit be to a direct current? How is this affected by short-circuiting the condenser? And what will be the admittance of the circuit under this condition?

CHAPTER III

PRODUCTION OF ELECTRICAL OSCILLATIONS AND RADIATION OF ELECTRICAL ENERGY

WE must now consider more fully what is meant by electrical oscillations.

If a weight on the end of a string is drawn to one side and released, it will swing backwards and forwards. The angle

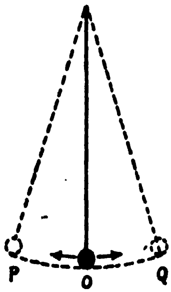


FIG. 17.

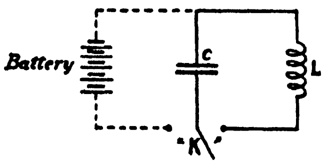


FIG. 18.

through which it swings will, by reason of the resistance of air, become smaller and smaller, so that finally it comes to rest.

Let POQ in Fig. 17 represent the path through which the weight swings during half a cycle. At P and Q the velocity of the weight is zero, but energy is stored up in it by reason of its position, i.e. it possesses *potential energy*. At O the weight is at the lowest point of the arc it is describing and possesses *kinetic energy*, i.e. energy due to its motion. This energy is sufficient to carry the weight to Q, where its velocity has disappeared and its kinetic energy has been reconverted into potential energy. At points between O and P or Q the energy of the weight is partly potential and partly kinetic.

The mechanical oscillations of a swinging weight are analogous to the electrical oscillations which take place in a wireless circuit. Instead, however, of mechanical kinetic and potential energy, we have electromagnetic and electrostatic energy.

Consider a circuit (Fig. 18) containing a condenser of capacity C and a coil of self-induction L, and suppose a key K is introduced into the circuit. Suppose the condenser is charged by

some means, for example by connexion to a battery. When the condenser is fully charged let K be thrown over. The condenser will then discharge through the coil and a current will begin to flow through the inductance L, starting from zero and gradually increasing to a maximum, when the condenser will be completely discharged and the potential between its plates will be zero. The current flowing through the coil will create a varying magnetic field in the coil and therefore produce an E.M.F. due to the self-induction of the coil. When the condenser is completely discharged, the current and the magnetic field through the coil will be at a maximum and will then begin to decrease. The effect of the self-induction of the coil will be to produce a gradually increasing potential which will charge up the condenser once more. The sign of the charges on the plates, however, will be reversed, and the plate which was charged positively to begin with will now be charged negatively. When the current has fallen to zero, the potential across the condenser will be a maximum and the condenser will again be charged. It will then begin to discharge once more through the coil, and after complete discharge the E.M.F. of self-induction will cause it to charge up again exactly as it was at the beginning of the cycle of operations. The whole cycle will then repeat itself.

If the potential across the condenser was V at the beginning of the cycle, we have seen in the last chapter that electrostatic energy equal to $\frac{1}{2}CV^2$ is stored in the condenser. When the condenser discharges, this energy reappears in the form of magnetic energy due to the current flowing in the circuit. If the maximum value of current is I, then, when the condenser is completely discharged, the magnetic energy will be equal to $\frac{1}{2}LI^2$. Thus, on the analogy of the oscillating weight, we get a continual transformation of electrostatic energy into electromagnetic energy, and we get what may be regarded as electrical oscillations taking place which manifest themselves as alternating currents and potentials in the circuit.

A circuit of this kind is known as a closed oscillating circuit. If no losses of any kind take place, the electrostatic energy is equal to the magnetic energy and therefore

$$\tfrac{1}{2}LI^2 = \tfrac{1}{2}CV^2.$$

If f is the frequency of the oscillations and $\omega = 2\pi f$, we have seen in Chapter II that $V = \omega LI$ or $V = I/C\omega$. Thus the fre-

quency at which oscillations take place, or the natural frequency of the circuit, is given by

$$\omega LI = \frac{I}{C\omega}$$

or

$$\omega = \frac{1}{\sqrt{(LC)}}$$

$$f = \frac{I}{2\pi\sqrt{(LC)}}.$$

By adjusting L and C we can cause the natural frequency of the circuit to have any value we please.

Hitherto the effect of resistance in the oscillating circuit has been disregarded, but no circuit is completely free from resistance. For, even if a separate resistance is not introduced, the connecting wires and the winding of the coil will have some resistance. Some of the energy which is released when the condenser discharges must be used in driving the current through the resistance of the circuit. This energy will appear as heat in the wires and will not be available as magnetic energy for recharging the condenser. Hence, when the condenser recharges, the energy restored will not be quite as great as it was originally. The action of the resistance of the circuit is analogous to that of the friction of the air on the oscillating weight. Owing to the energy lost as heat the electrical oscillations in the circuit will die away gradually, unless energy is continually added from outside by some means to maintain them. Thus the oscillating currents or E.M.F. produced will gradually decrease in their maximum values in each half-cycle, so that the oscillations can be graphically represented by a train, as shown in Fig. 2, p. 5. The rate at which the oscillations die away will depend on the amount of the resistance losses in the circuit. It may be remarked that, should the resistance be very high, no oscillations at all will take place, but the current from the condenser will increase to a maximum value and then leak away slowly through the resistance of the circuit. It can be shown mathematically that the condition for electrical oscillations to take place is that $R^2 < 4L/C$, and when the resistance is considerable, the complete formula for the frequency of the oscillations is

$$f = \frac{1}{2\pi}\sqrt{\left(\frac{1}{LC} - \frac{R^2}{4L^2}\right)}.$$

It will be seen that if $R^2/(4L^2)$ is small, this formula reduces to the expression already obtained for the natural frequency of the circuit.

It can also be shown that the train of waves given by the discharge of the condenser is such that the ratio of the successive maximum values of the current in the same direction is constant. The logarithm of this ratio is defined by a quantity δ, which is known as the 'logarithmic decrement', or more often as simply the decrement of the circuit, and is such that

$$\delta = \pi \frac{R}{\omega L}.$$

It can also be shown that the quotient of half the average energy dissipated per cycle divided by the average magnetic energy at the current maxima is equal to δ. For if I is the average effective current flowing in the circuit, RI^2 will be the average energy lost in the resistance per second, and if f is the frequency of the oscillations, the average energy lost per cycle will be RI^2/f. The square of the effective current in any cycle is equal to $\frac{1}{2}I_1^2$, where I_1 is the maximum current. The magnetic energy at the current maximum is equal to $\frac{1}{2}LI_1^2$, and therefore the average magnetic energy is equal to LI^2, so that

$$\frac{1}{2}\left(\frac{RI^2/f}{LI^2}\right) = \frac{R}{2fL} = \pi \frac{R}{\omega L} = \delta.$$

The losses of energy which cause the oscillations to die away are known as *damping* losses.

Hitherto we have only considered the energy lost as heat in the winding of the circuit itself, but any effect which consumes the energy of the condenser and prevents its reappearance as magnetic energy will produce a damping effect in the same way as ohmic resistance.

The most important causes of damping losses may be summarized as follows: (1) The ohmic resistance of the circuit; (2) currents induced into other circuits, or into masses of metal near at hand; (3) leakage through faulty insulation of the circuit; (4) energy radiated into the surrounding space. Each of these losses can be represented in the same form (namely RI^2) as the ohmic resistance losses, i.e. by a quantity equivalent to a resistance multiplied by the square of the effective current. Hence the expression for the decrement found above can be

applied to all these losses of energy. Each of these sources of loss is important when we are dealing with various aspects of transmitting and receiving apparatus, but the last requires special consideration, since it is by this means that the energy which conveys the radio-signals through space leaves the aerial. The transmitting aerial should therefore be designed so as to make this loss as great as possible.

Losses included under the heading (2) will repay further study, because, as we begin to employ shorter wavelengths or higher frequency oscillations, they become increasingly important. At very short wavelengths the inductance L will be small. It will contain few turns and so the ohmic resistance (1) of the circuit becomes almost negligible. Leakages due to (3) can be kept within reason by the use of good insulators having ample leakage path; whilst radiation resistance (4) is generally to be desired. Thus losses of type (2) remain as the most serious.

The induction of energy into other circuits or objects can best be avoided by their complete exclusion from the oscillatory field, and a coil carrying high-frequency current should be kept clear of surrounding objects as far as possible. Certain portions of the apparatus must be situated within the field however. These include the metal of which the coil L itself is constructed, the wire or tube with which it is wound, also any former on which the coil is wound and insulators upon which it is supported. The metal plates of the tuning condenser C will be in an intense field, and as a rule it is impossible to exclude all other components of a transmitter from the field, which may be quite extensive.

Now when any conductor cuts the lines of force of an alternating field, oscillatory currents will be induced in it. This effect is fundamental, and whilst it may be very useful in the design of a transformer, we cannot prevent it from happening at other less convenient times. Unwanted currents thus induced into a mass of metal are called 'Eddy Currents'. They always flow in such directions that the field they set up will oppose that which gave rise to them. In flowing through the resistance of the metal they cause heating, and thus waste energy from the main field. Eddy current losses can be reduced by breaking up the metal in which they flow, and inserting insulating layers in their path. This is done by the lamination of choke and

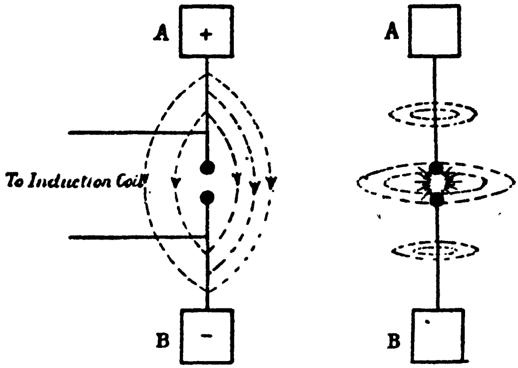
transformer cores, for example. In the case of high-frequency equipment the simplest remedy is to avoid excess metal in the field. Thus coils are built of tube or strip rather than solid wire. The high-frequency currents tend to travel in the surface skin of the metal only, whilst eddy currents flow through its inner thickness, which can be removed with advantage. A hollow conductor contains less metal in which these unwanted currents can flow.

There is one case in which eddy currents are useful. This is in the use of metal boxes or partitions to screen one high frequency circuit from the effect of another. Eddy currents set up in the partition will produce a field which opposes that causing them; and if this secondary field is of equal intensity it will reduce the resultant at outside points to zero. No resultant field then remains outside the partition to set up unwanted couplings. In this case we want to encourage the free flow of eddy currents within the screen, which should be a continuous conductor free from joints of high contact resistance. It should be of thick material having low internal resistance, such as copper. The induced currents will then be a maximum, but since they will encounter little resistance, there will be the least possible loss of energy as heat.

An effect similar to eddy current loss exists when imperfect insulators are within the oscillatory field. This is dielectric loss. Dielectric strain set up within the material causes heating, which may absorb considerable energy and also damage the insulator. In most materials this effect increases rapidly with frequency, until it may become the most serious of all at very short wavelengths. The heating of a poor material within an intense field may be sufficient to disrupt or set fire to it, whilst a massive coil former may introduce so much loss as to ruin the efficiency of a circuit. Therefore it is usual to employ the minimum possible amount of selected insulating material in the design of short-wave components.

The closed oscillating circuit which we have hitherto considered radiates very little energy into space. In order to increase the radiation it is necessary to increase the distance between the condenser plates, so that the capacity in the circuit, instead of being concentrated in the condenser, becomes distributed through the circuit. When this is done we have what is known as an open oscillator.

The simplest form of such an oscillator is the Hertzian oscillator described in Chapter I, page 2. Let the rods of such an oscillator be charged by an induction coil, so that A is positive and B is negative (Fig. 19 (a)). The region round the oscillator will now be in a state of strain which can be represented by lines of electric force between the charges on the rods. The effect is the same as in the case of the condenser in a closed oscillating circuit just before discharge.



(a) FIG. 19. (b)

When the charge on the rods grows sufficiently great, the strain between the spark balls will break down the resistance of the air between the balls and a spark will pass, and the resistance of the gap between the spark balls will become comparatively small. Then owing to the difference of potential between A and B, a current will begin to flow through the rods across the spark gap. As the current grows, an increasing magnetic field will be set up round the rods which will spread out in planes at right angles to the oscillator.

If the rate of discharge is comparatively slow, the electrostatic energy due to the electric strain will reappear as a magnetic field created round the rods. When the current across the spark gap has reached a maximum, the magnetic field will begin to die away and will cause the rods to be charged up again to the same potential. A will now have a negative and B a positive charge. The action so far is similar to that already described in the closed oscillator. The lines of magnetic force will collapse and those of electric strain will reappear in the opposite sense.

Suppose now that the discharge takes place very rapidly and, to make the matter as concrete as possible, let us consider two equal electric charges *a* and *b* in the rods and a line of electric strain connecting them. When the discharge begins, the charges *a* and *b* will move rapidly towards each other, and the lines of electric force joining them will contract as shown in Fig. 20 (*b*). The lines, however, may not contract as rapidly as the charges to which they are joined approach each other, so that we get the effect shown in Fig. 20 (*c*), where the ends of the

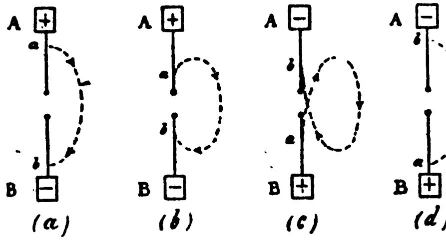


FIG. 20.

lines of strain have crossed and a loop of electrical strain in the surrounding medium is produced.

It is a well-known hypothesis of electrostatics that electrical lines of forces repel each other. Hence, the ring formed will be repelled from the oscillator while the electric lines of strain due to the charging of the oscillator in the opposite sense are being produced. Thus, when the oscillator is completely recharged, the ring will be pushed out into space as shown in Fig. 20 (*d*). Hence, as the oscillations continue, loops of electric strain will be repelled into the medium and the energy represented by them will not return to the oscillator, but will be radiated away in all directions round the oscillator.

We now come to the great contribution to the subject made by Clerk Maxwell. This is the hypothesis that a moving or varying electric strain is equivalent to an electric current as regards its magnetic effects. Maxwell called such a moving electric strain a *displacement current*.

The result of this hypothesis is that a moving or varying electric strain in the ether gives all the effects caused by a varying electric current or a moving charged electric conductor as regards the production of magnetic lines of force.

Now, a moving conductor in which a current is flowing will

produce a magnetic field at right angles to itself. Hence, the loops of electric strain repelled from the oscillator will produce magnetic fields at right angles to themselves and to the direction in which they are being propagated outwards, as shown in Fig. 21.

It must be clearly understood that the magnetic forces produced by the movement outwards of the electric strains are distinct from the magnetic fields produced by the actual currents in the oscillator rods. These latter magnetic fields are in phase

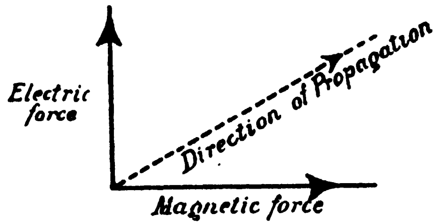


FIG. 21.

with the currents in the rods. The magnetic forces produced by the radiated electric strains, on the other hand, are in phase with the electric strains. The magnetic lines of force due to the current in the oscillator itself collapse into the oscillator when the current dies away. These magnetic forces are known as the induced field, and their effect will only be felt a short distance away from the oscillator. The magnetic fields, however, associated with the electric strains radiated away from the oscillator do not collapse on the oscillator rods, but spread to great distances.

The description which has been given of radiation is very incomplete. It is impossible, however, to evolve any complete picture of the way the radiation from a Hertzian oscillator or a wireless aerial actually takes place, or how these radiated disturbances develop electromagnetic waves in the ether. The reason for this is partly because nothing at present can be satisfactorily postulated about the ether, save that it is capable of transmitting electromagnetic waves with a definite velocity, and partly because the production of magnetic forces at any point in the ether is complicated by the fact that we are dealing with electric strains that are both moving and varying.

Starting from the hypothesis already stated, viz. that varying

electric strains produce the same effect as electric currents, Clerk Maxwell developed a series of mathematical equations connecting together the changing electric and magnetic forces in the ether. When these equations are examined in conjunction, it is found that equations of motion of the lines of electric and magnetic forces can be obtained which are exactly analogous to the well-known equations representing wave motion in other branches of physics.

Maxwell then formulated the hypothesis that radiated electric energy is propagated through the medium in the form of electromagnetic waves.

Evidence in support of the theory has been obtained from all branches of physics, but it must be remembered that Maxwell's theory is based on mathematical conceptions which in many cases cannot with our present knowledge be entirely translated into concrete physical realities. Some assistance can, however, be obtained from the consideration of wave motion in other media.

If we consider, for example, wave motion in a sheet of water, it will be found that each particle of the medium is moving upwards and downwards. Each particle moves in a straight line out from its position of rest a certain distance and then returns and moves out the same distance in the other direction. The time which the particle takes to complete its whole cycle is always the same.

It can be said, therefore, that the particle is executing a *periodic motion* in a straight line. All the particles in the medium will follow the same periodic motion, but at any instant the displacements of the particles (i.e. the distance of the particles from their original position of rest) are different, and the motion of the medium that takes place is, as it were, handed on from one particle to the next. If we watch the medium from a distance it will be seen that waves are being propagated with a definite velocity. If, on the other hand, we could fix our attention on one particle only, we should observe that the particle simply moves to and fro about a particular point.

Suppose we have a row of particles which were originally in the same horizontal layer of the medium in which the wave motion is taking place, and suppose the particle P (see Fig. 22) executes a simple periodic motion along YY_1 at right angles to this layer. Then if we follow the line XX_1 through this

layer, we come to a place where there is a particle P_1 which has exactly the same displacement O_1P_1 from XX_1, both in magnitude and direction, as the particle P. The distance OO_1 is called the wavelength of the disturbance and is usually denoted by the Greek letter λ. The time taken for P or any other particle to make a complete oscillation is called the periodic time of the disturbance. If this time be denoted by T, the velocity with which the wave appears to move forward along XX_1 is given by

$$Tv = \lambda.$$

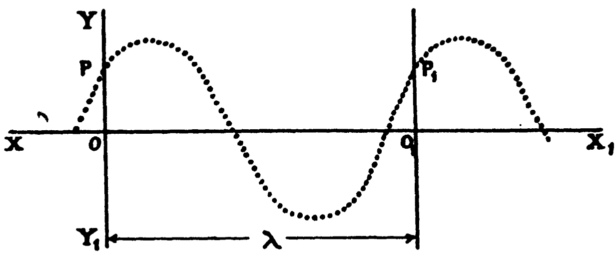


FIG. 22.

The number of oscillations per second made by a particle is known as the frequency f, and it follows that $f = 1/T$. The velocity with which electromagnetic waves travel through the ether is the velocity of light, viz. 3×10^{10} centimetres per second.

We have seen above that, in order to make the energy radiated as large as possible, it is necessary to use some type of open oscillator. In radio-telegraphy it has been found that one of the most efficient forms of oscillator is an elevated aerial in which wires are carried to as great a height as is convenient on masts or towers. In general, such an aerial can be considered as an oscillator, one-half of which is buried in the ground; and an earthed aerial of this kind is termed an open or Marconi aerial. It is employed in modern communication only for long wavelength working. An aerial is diagrammatically represented on the left of Fig. 23, and the loops of electric strain, due to successive current oscillations in the aerial, which are radiated out from it, are at some distance from the aerial represented by the lines of electric force in the figure.

The magnetic lines of force produced by the lines of strain moving forward will be perpendicular to the lines of electric strain and parallel to the surface of the earth. In the figure the

magnetic lines are represented thus, ⊕, if the direction of the magnetic force, which is at right angles to the page, is away from the reader, and thus, ⊙, if towards the reader. The magnitudes of the lines of magnetic force are roughly represented by the size of the circles. The whole groups of electric and magnetic forces are moving forward with the velocity of light, and the magnitudes at any point on the surface of the earth represent only their instantaneous values. The forces are

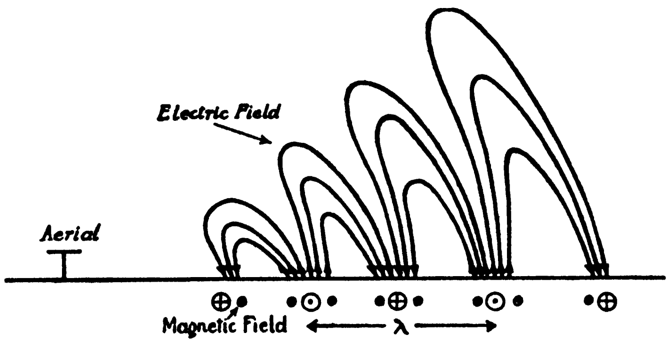


FIG. 23.

rapidly changing in an analogous manner to that in which the particles were moving in our illustration of wave motion in water. The distance from one point of instantaneous maximum intensity to the next of the same sign is equal to λ, the wavelength of the disturbance.

The intensity of radiation from the aerial at any point can be calculated either from the consideration of the electrical strains produced or from consideration of the magnetic forces caused by these strains.

We have considered the state of things at some distance from the aerial. Near the aerial, however, it is necessary to take into account the effect of the magnetic field due to induction produced by the currents in the aerial wire itself. In the induced field the electric and magnetic forces in the ether are almost exactly 90° out of phase. It can be shown that at any point at a distance d from the aerial the effect of the radiated field is proportional to $1/d$, while that of the induced field is proportional to $1/d^2$. At considerable distances attenuation may occur through such effects as absorption by the ground. Hence the field strength would be expressed by a function such as Yd^n;

n would then be the 'attenuation factor' for the wave. Thus the effect of the induced field dies away very rapidly compared to the radiated field, and at a distance of two or three wavelengths from the aerial it can be neglected altogether. Its existence is of importance, however, when it is a question of making measurements of the radiation close to an aerial. The field strength at any given distance from an aerial is expressed as the number of millivolts or microvolts oscillatory potential induced across a conductor one metre long. The units of radiated field strength are thus 'millivolts per metre'.

It is now necessary to consider how a Marconi type aerial may be made most efficient as a radiator of energy. The energy radiated is represented by an expression $R_a I^2$, where R_a is known as the radiation resistance, and is approximately equal to 1600 $\{(\alpha h)/\lambda\}^2$, where h is the geometrical height of the aerial, α is a factor depending on the design of the aerial, and λ is the wavelength of the wave emitted, which is assumed to be greater than h.

It is clear from the above formula that one of the chief requirements for efficient transmission is height. The masts, therefore, used at high-power long-wave stations are very high. Those, for example, at the Rugby Post-Office Station, which is capable of transmitting to Australia and New Zealand on long wavelengths, are 820 feet in height.

The reason for the introduction of the factor α is that the distribution of the current in the aerial is not uniform. In a simple vertical wire the current distribution is a maximum at the foot and decreases to zero at the top. The power radiated from any small portion of the aerial is proportional to the square of the amplitude of the current flowing in it. Each portion of the aerial being at a different distance from the ground has a different capacity to earth, and the capacity of the aerial is thus distributed. The distribution of current in the aerial depends on how its capacity is distributed. The factor αh is therefore frequently expressed as equal to the height of the *centre of capacity* of the aerial.

The height of the centre of capacity can be made greater by increasing the capacity of the upper portion of the aerial, e.g. by adding a horizontal portion to the aerial. The greater the number of wires forming the horizontal portion, the greater the increase in the capacity of the aerial. The value of α for

ordinary types of aerial may vary between 0·5 and 0·7. In many types of high-power aerials the capacity of the flat portion of the aerial is so great that the current in the vertical portion becomes practically uniform and approximately $\alpha = 1$, or the height of the centre of capacity becomes practically the same as the geometric height of the aerial. The effective height of the aerial, however, will be reduced by the capacity effects introduced by the masts and stays, especially if the masts be of metal. Very frequently the masts are insulated by the introduction of large insulators at their foot.

It must be remembered that, when large currents are to be induced in the aerial, the capacity of the aerial must be sufficiently great to prevent too large a voltage being developed; otherwise the voltage will cause the insulation of the aerial to break down, and 'brush' discharges will take place with corresponding loss of power.

It has already been stated in dealing with damping effects that the various losses, other than those due to radiation, due to resistance, &c., in an oscillating circuit, of which an aerial is an example, can be grouped together as an expression of the form RI^2.

If the power radiated is given by $R_a I^2$, the efficiency of the aerial as a radiator of power is given by

$$\frac{R_a}{R+R_a}.$$

It is thus clear that, in order to improve the efficiency of an aerial, R_a must be made as great as possible and R as small as possible.

As regards the latter factor, namely the reduction of R, several considerations present themselves. At one time it was thought necessary to provide a good earth connexion. This was usually obtained by arranging a number of earth plates connected by wires in the form of a fan under the aerial. The precise function of the earth connexion is somewhat obscure. One way of looking at the matter is that a system of earth plates makes the ground below the aerial a good conductor and enables the feet of the electrical lines of force to attach themselves immediately to the ground. Also, if the ground under the aerial is a good conductor, it can be shown mathematically that the elevated portion of the aerial forms an electrical image of itself

in the ground, so that the radiating portion of the aerial is in effect increased.

If an earth connexion, or 'buried earth', is employed, it is certainly necessary for efficient transmission that it should be a very good one. Later research, however, has shown that even better results may be obtained by the use of what is known as an *earth screen*. The earth screen, or 'capacity earth', consists of a network of wires beneath the aerial carried on insulators supported on posts, perhaps ten to fifteen feet above the ground. The action depends on the condenser effect of the screen, and it is therefore important that the screen should be carefully insulated. The remarkable results which can be obtained with such an arrangement are well illustrated by even some of the earliest experiments made with it.

H. J. Round[1] has published the following data concerning the results obtained with the aerial at the Clifden Station (now destroyed) of the Marconi Company. The average effective height of the aerial was 100 feet, and with original earth plates the total aerial resistance was 4·5 ohms. On a wavelength of 5,700 metres the radiation resistance was 0·05 ohm, and the radiation efficiency was about 1 per cent. When an earth screen was introduced the aerial resistance was reduced to 0·6 ohm, so that the radiation efficiency was increased to about 8 per cent. The aerial employed was very low: if its height had been 100 metres instead of 100 feet, the radiation efficiency would have been increased to nearly 40 per cent.

The earthed, open, or Marconi aerial has been described in some detail as it is still in frequent use for the transmission of waves longer than about 100 metres, for radio broadcasting, and for non-directional commercial services generally. It will be found as the main aerial on board ship or at the air ports, and is used by many mobile services because of its simplicity. A valuable property of this type of aerial is that it can be used over a wide range of wavelengths, the natural wavelength of the aerial being increased by the addition of inductance and capacity in the form of coils and condensers at the foot. The insertion of inductance between the foot of the aerial and the earth is known as 'loading', and a certain degree of loading is necessary in order that power may be introduced into the aerial by induction between the loading coil and a second coil

[1] 'Clifden'; H. J. Round, *Radio Review*, vol. ii, p. 459. 1921.

coupled to it. These properties of the earthed open aerial make it very suitable for general reception, when it is desired to pick up signals from any direction and at a variety of wavelengths; it is thus the type of aerial employed for domestic broadcast receiving and for many commercial services also.

Earthing the aerial was an expedient originally adopted by Marconi to give the effect of increased size, since, as has been explained, an electrical image of the aerial may be considered to be formed in the ground beneath it, and hence the size of the whole system is in effect doubled. When working with long waves this is very desirable, because the efficiency of the aerial increases with its height and size relative to the wavelength in use, and it is seldom possible to erect as large an aerial as might be wished. At the time, earthing the aerial produced a great improvement over earlier methods and was a decided step forward.

The original Hertzian oscillator of Fig. 19 through which we approached the idea of radiation from an aerial was not earthed, however, but consisted of two symmetrical conductors terminating in plates for the purpose of increased electrical capacity. This oscillator is an even more effective radiator than the earthed aerial, provided that the distance from *A* to *B* is of the order of half the transmitted wavelength, and therefore when short wavelengths are to be used it becomes possible to dispense with the earth connexion entirely, and to use a simple oscillator of sufficient size. When the distance *A* to *B* is many metres it is found unnecessary to retain the capacity plates *A* and *B* themselves, since the rods alone have sufficient capacity, and thus the aerial becomes simply two equal rods or wires of total length nearly half a wavelength, suspended in free space. The salient characteristic of Hertzian aerials is that the whole aerial system is complete without any form of earth connexion, and the capacity of the system to earth plays no essential part in its operation.

With the gradual adoption of short-wave communication in recent years the Hertzian type of aerial has gradually superseded the Marconi, until at the present time it is probably the more important of the two. The greater part of radio-communication now takes place on wavelengths below 100 metres, and in this region the Hertzian aerial is both practicable in size and the most efficient radiator.

The radiation from a Hertzian aerial at short wavelengths is dependent upon the radiation resistance just as in the case of the earthed aerial. In fact, the expressions involving this were derived from a study of the simple oscillator in the earlier pages of this chapter. The other circuit losses differ somewhat, however, owing to the changing properties of the shorter waves, and the calculation of the radiation resistance is a little less simple. The very high-frequency oscillations which correspond to short wavelengths are more easily dissipated in the form of heat, and the resistance of conductors to them is greater than at lower frequencies. The various losses grouped together under the expression of the form RI^2, therefore, tend to become greater at the shorter wavelengths for which the Hertzian aerial is used; but this increase is usually more than compensated for by the lower resistance possible in an aerial of smaller dimensions, and by the avoidance of earth resistance losses.

The element of height enters into the efficiency of the Hertzian aerial in a different manner than into that of the earthed aerial. It is still important that the height be considerable, but there will be a maximum height beyond which little increased radiation results. The radiation is not equal in all directions, and the height above ground has an effect upon this factor which is often of major importance. Unlike the earthed aerial which must be at least partly vertical, the Hertzian type may be suspended horizontally or vertically at will, and its actual size and efficiency as a radiator is not directly determined by the height above ground. Height is valuable in this case to bring the aerial above the level of surrounding objects, which might absorb or intercept the radiation, and to raise it sufficiently far above the ground and buildings to minimize losses caused by induced currents in conducting objects below. Little is gained in most cases by raising the aerial more than one wavelength above ground, and this results in lower masts and less costly aerial systems than those necessary for effective long-wave working.

In this chapter we have dealt with the basic idea of the radiating aerial and with its use as a transmitter. Later it will be shown that the same considerations apply in the receiving aerial system, and the principles governing the design of modern short-wave aerial systems will be considered more fully. Before this

is attempted, however, it will be helpful to study the associated radio equipment so that the requirements of aerial systems can be more easily grasped.

EXAMINATION QUESTIONS

1. What is the impedance at a frequency of 900 kilocycles per second of a circuit consisting of a condenser of 0·0002 microfarad capacitance in series with a coil of 200 microhenrys inductance and 25 ohms resistance?

City and Guilds of London Institute. Preliminary Exam. 1934.

2. An antenna has a capacitance of 300 $\mu\mu F$ and an inductance of 30 μH. What series inductance must be added to tune it to a wavelength of 1,500 metres? What alternative method can be used to tune the aerial? *C. and G. of L. I. Preliminary Exam. 1935.*

3. What is meant by the 'radiation resistance' of an aerial? In the case of a given aerial how does it vary with the frequency to which the aerial is tuned? (See also chapter 14.)

Institute of Wireless Technology. May 1935.

4. Explain the difference in operation between an open or Marconi aerial and a Hertzian aerial. For what wavelengths would each be preferable, and for what reasons?

5. What is meant by field strength? The field strength due to a distant transmitting station is found to be 33·6 millivolts per metre at a distance of 100 kilometres, and 6·1 millivolts per metre at a distance of 300 kilometres. Calculate the attenuation factor of the wave. *I. W. T. June 1937.*

6. An inductance coil is connected across a standard variable condenser and tuned to the frequencies given below. If the added capacities required to tune the coil to these frequencies are as stated, calculate the value of the inductance and also the internal capacity of the coil.

Frequency:	1,000	750	600	500 Kc.
Added capacity:	0	70	160	270 micromicrofarads.

A. M. I. W. T. June 1937.

7. A condenser of 400 mmfd. capacity and negligible series resistance is connected across an inductance of 200 microhenrys which has an effective resistance of 10 ohms. Calculate the resonant frequency of the combination and also the voltage across the condenser at that frequency, if the current in the supply circuit has an R.M.S. value of 5 mA. *A. I. W. T. June 1937.*

8. An antenna has an effective capacity of 200 micromicrofarads and negligible inductance. In series with it is connected a condenser of 800 micromicrofarads and an inductance of 250 microhenries. What is the wavelength to which it is tuned? If the condenser be short circuited, what additional inductance must be added to tune the aerial to 942 metres? *Grad. I. E. E. 1936.*

9. At a frequency of 796 cycles per second, an E.M.F. of 6 volts sends 100 mA. through a certain circuit. When the frequency is raised to 2,866 cycles per second, the same voltage sends only 50 mA. through the same circuit. Of what does the circuit consist?

Grad. I. E. E. 1936.

10. What are meant by 'Eddy current losses' in an alternating-current circuit? What steps would you take to reduce these to a minimum in practice? Compare the effects of eddy current and dielectric losses as the frequency of an oscillatory circuit is raised from 10 kilocycles to 10 megacycles.

11. Explain the advantages of height in an aerial system. Are these identical in the case of Marconi or Hertzian aerials, and what is the reason for earthing the former?

CHAPTER IV

TRANSMISSION OF DAMPED WAVES AND THE RESONANCE OF TUNED CIRCUITS

THE various types of transmitting systems have already been referred to in Chapter I. We have seen that they can be divided into two main classes: (1) systems in which damped waves are emitted; (2) systems in which continuous waves are emitted.

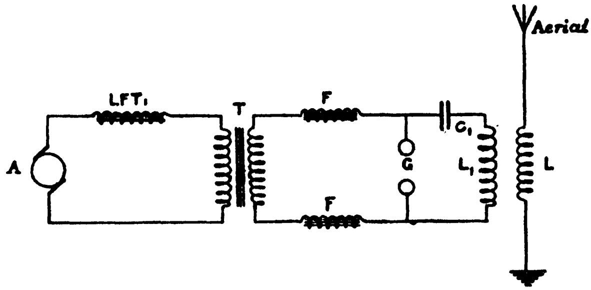


FIG. 24. Spark Transmitter.

A, *Alternating-current Generator.* T, *Iron-cored Transformer.* FF, *Air-cored Chokes.* LFT_1, *Low-frequency Tuning Inductance.* G, *Spark Gap.*

In the case of damped waves a dot or dash of the Morse code consists of a number of groups of high-frequency oscillations of continually decreasing amplitude. Such waves were largely used for ship installations. They are not so pure as continuous waves and the tuning is not so sharp. This may be an advantage in ship emergency installations, since of necessity distress signals must be easily received. When traffic is to be passed, however, damped waves have given place to continuous waves in all cases.

In this chapter we shall first consider the fundamental principles underlying the transmission of damped waves, because these form a convenient theoretical introduction and have historical interest.

We have already seen that the resistance of the air in a spark gap is practically infinite, but that when a spark takes place across it its resistance drops to a low value (one or two ohms). In Fig. 24 the condenser C_1 is charged up by means of an induction coil or some form of step-up transformer (i.e. an apparatus

by which the low voltage produced by an alternating-current generator, say 75 volts, is transformed to a high voltage of the order of, say, 15,000 volts). When C_1 is charged to a sufficiently great potential, the resistance of the air in the spark gap G breaks down and oscillations take place in the circuit $L_1 C_1$. The amplitude of the oscillations decreases, owing to the damping losses which take place, until after a few oscillations the potential of the condenser is insufficient to maintain the spark across the spark gap, and the resistance of the gap again increases. The condenser is then recharged by the transformer and another group of oscillations is produced.

Thus in the discharge of a condenser across a spark gap we have what amounts to an automatic switch replacing the key through which the condenser was discharged in Fig. 18. The coil L_1 in the condenser circuit is brought near a coil L in the aerial circuit, and the oscillations in the condenser circuit excite oscillations in the aerial circuit by the mutual induction due to the coupling between these coils.

We have seen that the frequency of the oscillations produced in the condenser circuit, neglecting the resistance of the circuit, will be given by the relation $f = \frac{1}{2\pi\sqrt{(L_1 C_1)}}$, where L_1 and C_1 are the self-induction of the coil and the capacity of the condenser measured in henries and farads respectively.

The aerial is also an oscillating circuit possessing a capacity which we may express by C and a certain self-induction expressed by L. It may be remarked that L is the sum of the self-induction due to the wire of the aerial and the self-induction of any coil introduced into the aerial. The aerial circuit has therefore a natural frequency of its own given by $f = \frac{1}{2\pi\sqrt{(LC)}}$. If we wish to excite in the aerial oscillations of the greatest possible magnitude with the power available, the natural frequency of the aerial circuit must be made equal to that of the condenser circuit. In other words, the coil in the aerial circuit must be adjusted until $LC = L_1 C_1$. This process is called *tuning*, and is fundamental to all modern systems also.

In actual working, the units we have hitherto used, namely henries and farads, are too large to express conveniently the comparatively small capacities and inductances used in most radio circuits. Instead of these units it is usual to employ as

units the microhenry, which is equal to one-millionth (or 10^{-6}) of a henry, and the microfarad, equal to one-millionth (or 10^{-6}) of a farad.

Also, instead of referring to the natural frequency of a circuit, it is a common practice to speak of the circuit as being adjusted to a given wavelength. Both forms of description are often used.

We have seen that wavelength is connected with frequency by the relation $f\lambda = V$, where V is the velocity of light, namely 3×10^8 metres per second. Thus, if the wavelength λ is measured in metres, then, since $f = \frac{1}{2\pi\sqrt{(LC)}}$,

$$\lambda = 3 \times 10^8 \times 2\pi\sqrt{(L_{\text{mics}} \times 10^{-6}\, C_{\mu F} \times 10^{-6})} = 1885\sqrt{(LC)}$$

where L and C are expressed in microhenries (mics) and microfarads (μF).

If we wish to adjust the transmitting set to emit a wave of, say, 600 metres, or, what is the same thing, oscillations of a frequency $\left(\text{given by } f = \frac{3 \times 10^8}{\lambda}\right)$ of 500,000 cycles per second, it is first necessary to choose or adjust L_1 and C_1 in the condenser circuit so that $\lambda = 1885\sqrt{(L_{1\,\text{mics}} C_{1\,\mu F})}$, where $\lambda = 600$ metres.

It is then necessary to tune the aerial circuit, including any coils and condensers introduced into it, also to 600 metres, i.e. to a frequency of 500,000 cycles per second.

As already stated, the Marconi aerial itself without the coupling coil has a certain capacity and self-induction. Two cases may arise: (1) the natural frequency of the aerial with a suitable coupling coil may be higher than that of the condenser circuit, i.e. the wavelength may be too low; or (2) the natural frequency may be lower, i.e. the wavelength may be too high. In the first case additional self-induction must be introduced by means of a coil whose self-induction can be adjusted by tapping off the correct number of turns until the self-induction of the aerial has the correct value. In the second case a condenser can be introduced into the down lead of the aerial circuit. Such a condenser, it will easily be seen, will be in *series* with the fundamental capacity of the aerial, and will have therefore the effect of decreasing the capacity of the aerial circuit.

Transmitting condensers usually consist of a number of thin metal plates connected in parallel and immersed in insulating

oil or air-spaced, and the capacity of the condenser is adjusted by altering the number of plates in it until the natural frequency of the aerial circuit is brought down to the correct value. These adjustments are carried out in practice with the help of an instrument called a wavemeter, which will be described later.

At first sight it might seem that more power could be got into the aerial by directly introducing the spark gap into the aerial circuit. This was indeed done in the early days of wireless. It will be seen at once, however, that if this is done the resistance of the spark gap is included in the aerial circuit and the oscillations produced across the spark gap are very highly damped. Thus, when a spark takes place, we get a large oscillation followed by only one or two others of very rapidly decreasing amplitude. An oscillation of this description is such as to set every receiving aerial within range in oscillation independently of whether it is tuned to the wavelength emitted or not. With such an arrangement, interference between neighbouring stations becomes so great that satisfactory communication is impossible. The arrangement has, therefore, been forbidden by international agreement except for distress signals, and is not now used even for this purpose.

For satisfactory working it is necessary that as much power as possible should be radiated on the particular frequency or wavelength allotted to the station, and as little power as possible upon other frequencies. This can only be done by careful tuning, so that the aerial and condenser circuits are in a state of resonance. A reference to Fig. 25 will show at once that this is impossible unless the resistance of the circuit is kept low.

Further, the degree of coupling, i.e. the tightness or looseness of the coupling between the coils of the condenser circuit and the aerial circuit, has a very important effect. The complete discussion of the effects of coupling between two currents in which high-frequency oscillations are taking place involves a very considerable knowledge of mathematics, but the general principles are fairly easy to understand.

Suppose we have two circuits A and B, each containing inductance and capacity, and suppose that oscillations are produced in circuit A (see Fig. 26) and that the coils $L_2 L_1$ are brought near together so that the magnetic field of L_1 links L_2. We shall refer to A as the primary circuit and to B as the secondary circuit. In the case of a spark-transmitter circuit,

A corresponds to the condenser circuit and *B* to the aerial circuit.

By the mutual induction between the coils L_1 and L_2 some of the energy of the oscillations in *A* is transferred to *B*, so that

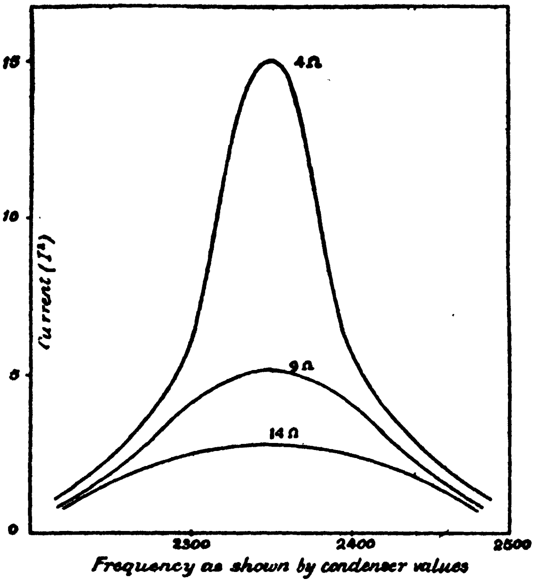


FIG. 25. Resonance Curves with Various Resistances, 4, 9, and 14 ohms, in Circuit.

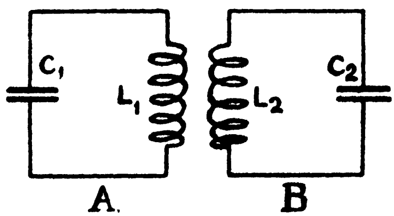


FIG. 26.

oscillations are thereby built up in *B*. This loss of energy from *A* will produce a damping effect in that circuit. Thus the oscillations in *A* will decay, while those in *B* will increase.

The transference from *A* to *B* will continue until all the available energy in *A* has been transferred to *B*. When this has taken place the oscillations in circuit *B* will influence circuit *A* and the energy of the oscillations in *B* will be transferred back

again to A, and the process will repeat itself, the energy being handed backwards and forwards from one circuit to the other.

The losses due to resistance and radiation, which we have so far neglected, are continually decreasing the total amount of energy available during each exchange, and the oscillations in both circuits will finally die away. With an aerial circuit coupled with a circuit containing a spark gap the resistance of the spark gap introduces a considerable damping.

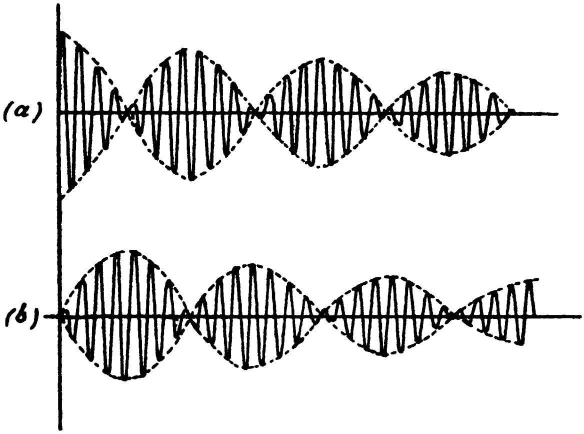


FIG. 27. Tight Coupling.
(a) *Oscillations in Circuit* A. (b) *Oscillations in Circuit* B.

If the coupling between the circuits A and B is tight, the energy will be transferred rapidly backwards and forwards between the circuits, as shown in Fig. 27.

If we draw the dotted line through the maximum value of the successive oscillations, we see that the effect of the tight coupling is to produce a slow oscillation in the circuit represented by the dotted line. The effect is analogous to the production of 'beats' in sound.

If we separate the two coils of the circuits A and B, i.e. if we loosen the coupling, the transfer of energy between the circuits will be much slower, as shown in Fig. 28, and therefore the beat effect will be correspondingly less pronounced.

Now we know that, in sound, beats are produced by the interference of two waves of different frequency, and in the present case, when we examine mathematically what is taking place in the coupled circuits, we find that two oscillations of different frequency are actually being produced.

Now if the circuits A and B are tuned to resonance, then the product L_1C_1 must equal L_2C_2. If the coupling is so close that all the lines of force produced in L_1 link L_2, the mutual induction M between the circuits can be shown to be such that

$$M = \sqrt{(L_1L_2)}.$$

If, however, the coupling is not so perfect, we can express the degree of coupling by a factor k such that $M = k\sqrt{(L_1L_2)}$. For

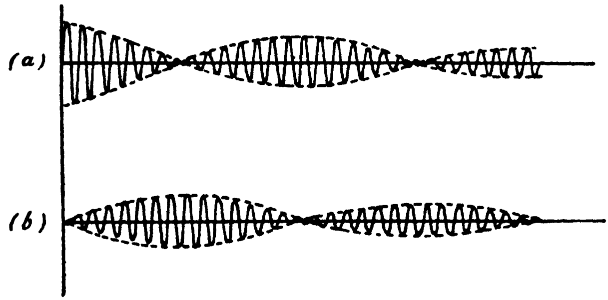


FIG. 28. Loose Coupling.
(a) *Oscillations in Circuit* A. (b) *Oscillations in Circuit* B.

a perfectly tight coupling $k = 1$, and for a perfectly loose coupling, where no lines from L_1 link L_2, $k = 0$.

In intermediate cases, instead of A oscillating at one frequency, deduced from the product L_1C_1, it will oscillate at two frequencies derived from $L_1(1+k)C_1$ and $L_1(1-k)C_1$ respectively.

Similarly, two frequencies will be produced in circuit B, which can be calculated from the products $L_2(1+k)C_2$ and $L_2(1-k)C_2$. Thus, in the case of an aerial circuit, if the mean frequency f is given by $f = \frac{1}{2\pi\sqrt{(LC)}}$, two waves will be emitted corresponding to

$$f_1 = \frac{f}{\sqrt{(1+k)}} \quad \text{and} \quad f_2 = \frac{f}{\sqrt{(1-k)}}.$$

The length of the waves will be given by

$$\lambda_1 = \lambda\sqrt{(1+k)}, \qquad \lambda_2 = \lambda\sqrt{(1-k)}.$$

If we draw a resonance curve representing the radiated energy plotted against the frequency, it will have two peaks

representing the two oscillations, as shown in Fig. 29 (*a*). If the coupling is reduced, these peaks come closer together, and if the coupling is very loose they practically coalesce and we get a single peak, as represented in Fig. 29 (*b*).

The coupling between the aerial circuit and the condenser circuit will depend on the closeness of the coupling coils and on the number of turns of wire in the aerial coil. Since the

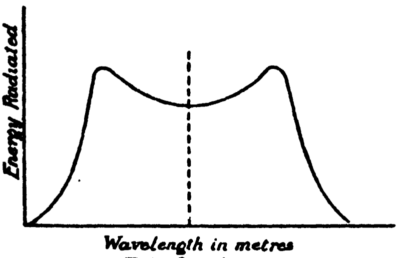


FIG. 29 (*a*).

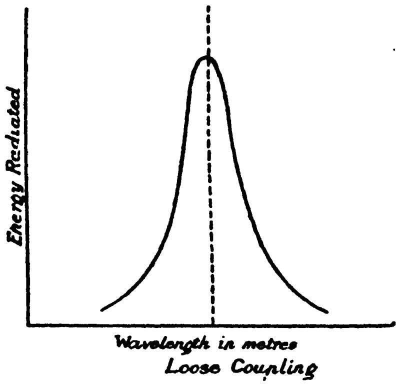


FIG. 29 (*b*).

coupling is proportional to $M/\sqrt{(L_1 L_2)}$, it is clear that the coupling will be less in proportion as L_1 and L_2 are greater if M is constant.

The general question of tight or loose coupling has been explained somewhat fully in its relation to spark transmission because, although spark is now practically obsolete, it provides a simple approach to the theory of coupled circuits, which are of equal importance in modern equipment also. It will be convenient at this point to follow the question a little further and, setting aside for the moment spark transmission, to look at coupled circuits from the receiving point of view.

Consider firstly the case of loose coupling illustrated in Fig. 29 (*b*). Here the resonance curve is single-peaked at one frequency only, and resembles that of a single circuit shown in Fig. 25 except that the slope of the curve is somewhat steeper. We shall have occasion to refer very frequently to the curves of Fig. 25, and to other very similar resonance curves, since they form the most convenient way of representing resonant circuit conditions. The curve obtained by the very loose coupling of two similar circuits is therefore of the same general shape

as that of either circuit taken separately, but the slope of the curve will be steeper. The combination may be said to be more 'sharply tuned' than were the original single circuits. Now, it is seen that the resonance curve of the single circuit of Fig. 25 becomes sharper as the circuit resistance is reduced, or as the total damping from all causes becomes less. It might be thought that the curve could be sharpened indefinitely if the resistance were sufficiently reduced. This, however, is unfortunately not the case, for not only is it impossible to make a coil of zero resistance, but also, as in the case of an aerial, the resistance term is in practice a complex expression which contains any losses due to imperfect materials from which the coil and condenser are built, or to surrounding objects in which currents may be induced. Additional damping may arise from the load imposed on the circuit by others to which it must be connected, and the losses in which can be compared to an additional resistance inserted into the resonant circuit itself. When, therefore, we need a sharper resonance curve than that of the best obtainable single circuit, it can be obtained by loosely coupling two such circuits and treating the result as a unit. It is even possible to couple three or more circuits in this manner, thereby obtaining a very sharp resonance curve indeed. If the coupling be magnetic, or of any form which can transfer energy with equal ease in both directions, the resulting sharpened curve differs more and more from that of a single circuit, approaching more nearly the shape of a narrow rectangle. Where a form of one-way coupling is resorted to, such as through the medium of the thermionic valve, the shape of the final curve is obtained by raising the ordinates of the single-circuit resonance curve to the nth power, where n is the total number of coupled circuits.

For two circuits to be coupled it is not essential that they be magnetically linked or coupled inductively. Any method whereby energy is transferred from the one to the other will effect coupling, and the form of the resulting curves will be similar. The introduction of an impedance common to both circuits does this, and circuits which are magnetically isolated can be coupled by either a resistance, an inductance, a condenser, or a mixed impedance Z common to both circuits, in the way shown in Fig. 30. This form of coupling is often more convenient in modern circuit construction than is direct magnetic coupling. It enables the two circuits to be isolated in

screened compartments from all outside fields, or to be any convenient distance from each other; whilst the tightness of coupling can be controlled by the variation of the impedance Z. Another useful property is that when desired the coupling can be made to vary with changing wavelength by the selection of the reactance Z. If this be a pure inductance its reactance will rise with reduced wavelength (increased frequency) and the coupling will become tighter; whilst if Z be a capacity its reactance will fall with increasing frequency and the coupling will be reduced. By the use of a combination of inductance and capacity, or of a pure resistance, the coupling can be caused to change in any desired manner with frequency, or to remain independent of it. This factor is often an assistance in circuit design.

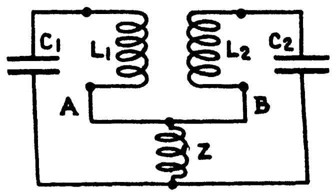


FIG. 30.

The radiation of energy by an aerial is a reversible process, and similar although much smaller oscillatory potentials to those in the transmitting aerial will be induced into a similar receiving aerial exposed to the radiated field. This matter will be more fully explained when we come to deal with reception. In exactly the same way the resonance curves which we have been discussing will remain identical if the resonant circuits form part of or are coupled to a receiving aerial. Thus if we read 'Energy Received' in place of 'Energy Radiated' in all the curves given, they will represent at once the ability of these circuits to separate the signals to which they are tuned from others on neighbouring wavelengths. It is for this purpose that resonant circuits are used in reception, to discriminate between the wanted signals to which they are tuned and to which the circuits respond most readily, and those at other wavelengths. The relative response to these will be proportional to the ordinates of the curves at the wavelength of the wanted and interfering station, and for this to be a minimum a sharp resonance curve is obviously necessary. It is for this reason that the effect of a series of coupled circuits is important, and their use will be explained more fully in the chapter on selectivity.

From Fig. 29 (*a*) we see that excessive coupling results in a double-peaked response curve. In reception the feature of this

will be that of fairly uniform response over a band of wavelengths (or frequencies) equal to the spacing between these peaks. This can be varied by different degrees of coupling, and hence in the tightly coupled circuits of early spark transmission we find the original of a popular modern device, the 'band-pass filter'. It will not require added explanation to see that a pair of tightly coupled circuits will respond to a band of frequencies of equal width on each side of the actual resonant frequencies of the individual circuits, whilst having little or no response to all others. The title band-pass filter is therefore easily understood. Under certain circumstances it is desirable to receive a band of frequencies on each side of the actual incoming wavelength, and when this is the case the closely coupled circuits have been found to provide a very useful response curve.

Throughout the preceding paragraphs the terms frequency and wavelength have been freely interchanged when speaking of a radio wave, and it must be clearly grasped that they are merely different aspects of the same physical reality. It is unavoidable that one or the other of these be employed at intervals through the book, that chosen being the most appropriate to the subject dealt with. It is most usual to-day to talk in terms of frequency, and we shall do this in most cases. When discussing aerials, however, wavelength will be more easily visualized.

Returning to the generation of damped waves, we see at once that the best resonating effects can be excited in the aerial circuit when this circuit is unable to react back on the condenser circuit.

To achieve this it is obviously desirable to arrange by some means that, when all the energy in the primary circuit has been transferred to the aerial circuit, the spark in the primary circuit is extinguished, so that the spark-gap resistance becomes very great and the circuit containing it is unable to oscillate. If this is done, the aerial circuit will go on oscillating at its natural frequency and one wave only will be emitted. This action is called *quenching*. The two methods most frequently employed for quenching were either the use of a rotating spark gap or a specially designed spark gap known as the 'quenched gap'.

In the rotating-gap method, instead of using two fixed spark balls, the spark is produced between projecting studs on a rapidly revolving metal disk and two fixed electrodes. The studs

and the fixed electrodes are adjusted so that the clearance is as small as possible. A spark will take place when the condenser voltage is sufficiently large to break down the insulation of the air in the spark gap, and the path of the oscillating current will be through one electrode, across the rotating disk, and out through the other electrode. Thus there are in fact two spark gaps in series. The rotation of the disk will rapidly increase the distance between the studs and the electrodes, so that the spark is quenched and the oscillation in the primary circuit ceases.

To hasten further the extinction of the spark, it was usual in large sets to direct a blast of air on the spark gap by means of an electric blower.

In the 'quenched-spark' system developed by the German Telefunken Company, the effect of quenching was produced by rapidly cooling the path taken by the spark. To do this, use was made of the property of metals for conducting heat. Instead of employing a single spark gap produced between spark balls, the discharge was produced between a series of large and well-cooled copper surfaces spaced about 0·2 mm. apart and separated by thin rings of mica. The mica rings were introduced in order to shut in the spark gaps from the outside atmosphere, as the sudden rise of pressure produced when the discharge took place in the small volume of air between the copper plates was found to add materially to the rate of quenching.

In all damped-wave systems in which the input power was derived from an alternator, it was usual to insert a variable iron-cored inductance in the primary of the transformer circuit (see LFT_1 in Fig. 24). The object of this was to enable the primary and secondary circuits of the transformer to be tuned to resonance with the low-frequency alternating current supplied by the alternator. Also, in order to prevent the high-frequency currents developed in the condenser circuit from coming back into the transformer windings, air-cored inductance coils or choke coils were inserted in the leads from the transformer to the spark gap (see Fig. 24).

Finally, a sending key is necessary by which the dots and dashes of the Morse code can be produced. In small-power sets this is inserted directly in the low-voltage side of the transformer. In large-power sets, of 5 Kw. input and upwards, a magnetic key or keying relay may be used, inserted in the high-tension side of the transformer. In this case the actual sending

key is placed in an auxiliary circuit supplied by direct current. When the key is pressed the current in the auxiliary circuit passes through a bobbin, which causes a magnetic force to act on a steel rod, passing through the bobbin, to which a copper arm is attached. The copper arm is thus caused to move and to make contact in the transformer circuit. The action will be clearly seen by reference to Fig. 31. To cause the making and breaking of the circuit by the copper arm to be rapid and clean,

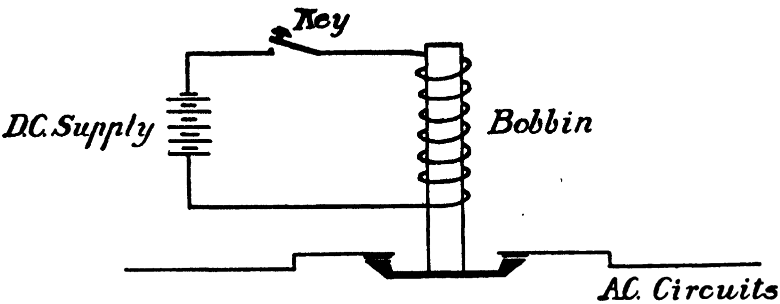


FIG. 31.

a blast of air from a blower is directed on to the key contacts in the alternating-current circuit. There are, of course, very many different designs of relay switches, but their action is in principle that described above. The relay principle in general is of importance in modern radio systems, since it is also used to enable the high-power circuits of large transmitters to be controlled from a central control desk. Relay keying is used also in continuous-wave transmission whenever it is necessary to break large currents, and in all those cases, now very common, where a radio-transmitter is keyed automatically by incoming signals over a telegraph cable. There is a tendency in modern equipment to key whenever possible at points in the circuit where the power is small, and keying systems will be further explained from time to time, as the need arises.

EXAMINATION QUESTIONS

1. What is meant by a damped oscillation? How are such oscillations produced? Why is the use of damped oscillations for radio-communication purposes declining?

City and Guilds of London Institute. Preliminary Exam. 1936.

2. In a spark transmitter the closed-circuit condenser has a capacitance of 0·16 microfarad. The high-tension power transformer has a step-up ratio of 1 : 30. What inductance will be required in the primary circuit for low-frequency tuning, if the frequency of the supply is 200 cycles per second?

C. and G. of L. I. Preliminary Exam. 1937.

3. In a spark transmitter, what is the effect on the emitted waves of increasing the coupling between the primary circuit and the aerial? What is the difference in this respect between a quenched-spark and an unquenched-spark transmitter?

C. and G. of L. I. Preliminary Exam. 1933.

4. Two circuits, each consisting of an inductance of 2 millihenries and a resistance of 200 ohms and a condenser of 0·002 microfarad, are coupled by a mutual inductance of 10 microhenries. At what frequency will an E.M.F. of one volt in one of the circuits produce maximum current in the other? What will be the values of the current in each circuit? *Grad. I. E. E. 1936.*

5. Compare spark, interrupted continuous waves, and pure continuous waves as methods of communication. What are their relative merits in point-to-point communication services, and for emergency working at sea?

6. In a spark transmitter the primary condenser has a capacitance of 0·016 microfarad and is charged from an alternator through a transformer having a step-up ratio of 100. If the inductance of the alternator is 0·05 henry and the frequency of the alternator is 50 cycles per sec., what additional inductance should be added to the low-tension circuit to produce resonance?

C. and G. of L. I. Grade 1. 1933.

7. Explain how two oscillatory circuits may be coupled together through the effect of a common impedance. Mention advantages of this procedure. How would you expect the coupling to vary with frequency if the common impedance were a condenser, an inductance, or a pure resistance?

8. What is meant by a band-pass filter, as applied to radio-receiver design? Outline the effect upon the shape of the overall-resonance curve obtained by the use of a number of tuned circuits in cascade.

CHAPTER V

THERMIONIC VALVES AND THE CATHODE-RAY TUBE

BEFORE dealing with the action of thermionic valves, it will be necessary to recapitulate briefly a few facts concerning the modern views of electricity and matter. A body conducts electricity because it contains a vast number of free electrons. A cubic centimetre of copper, for example, has been estimated to contain 10^{24} free electrons, or somewhere about double the number of molecules present in the same volume. The charge of negative electricity carried by each electron is about $1 \cdot 56 \times 10^{-19}$ coulomb, and its mass is about $8 \cdot 8 \times 10^{-28}$ gramme. The diameter of a hydrogen atom is about $2 \cdot 17 \times 10^{-8}$ cm., and an atom is from 50,000 to 60,000 times larger than an electron. If an atom were magnified until it was about the size of the dome of St. Paul's Cathedral, an electron would be of the order of a pin's head.

The atoms composing a body are made up of a positive nucleus round which revolve various numbers of electrons, making, as it were, miniature solar systems. Thus a free electron may be supposed to pass through an atom without as a rule disturbing the bound electrons revolving round the positive nucleus, much in the same way as a comet can pass through our solar system. The motion of these free electrons under an applied difference of potential causes what is known as an electric current.

We know that when a liquid is heated evaporation occurs. On the kinetic theory this phenomenon can be explained as follows. The molecules are in a state of motion, and under the influence of heat the velocity with which they are moving increases, until it becomes sufficiently large to carry the molecules outside the sphere of attraction of the surface of the liquid, so that they escape into the surrounding atmosphere. In the same way, when a conductor is heated the velocity of the free electrons is increased. Some of the electrons may be driven out from the surface of the conductor, and a phenomenon similar to evaporation occurs. In this way a stream of electrons may be emitted from an incandescent wire, and an electronic emission current or, as it is often called, a thermionic current

can be obtained. The nature of this emission was first definitely shown by J. J. Thompson in 1899, when by measuring the ratio of the charge on each particle to its mass he showed the particles to be electrons.

In 1903 it was pointed out by Professor O. W. Richardson that a mathematical expression, of the same form as that known to connect the vapour pressure and temperature in ordinary evaporation, could be applied with suitable modifications to the similar phenomenon of electrical evaporation. On this theory the electronic currents from various substances have been investigated by Richardson and others. Langmuir, for example, has given curves for the electronic current emitted from tungsten in a perfect vacuum, which show the electron emission as 4 milliamperes per sq. cm. at 2,000° K., about 60 milliamperes at 2,200°, 400 milliamperes at 2,400°, and 2,000 milliamperes at 2,500°. The current is different for different substances at the same temperature. The action of the thermionic valve is based on this phenomenon of electronic evaporation.

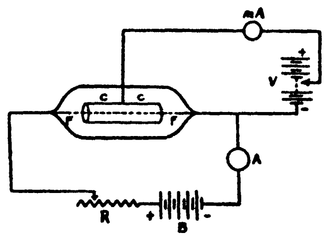


FIG. 32.

Suppose we have a thin tungsten wire filament sealed into a glass vessel, similar to an electric-light bulb, which is exhausted by a system of air-pumps until a very high vacuum is obtained, and suppose that this wire is surrounded by a metal cylinder and that connecting leads attached to the filament and metal cylinder are brought through the glass vessel by suitable air-tight seals. The filament can be made incandescent and heated to a temperature of, say, 2,400° K. by connecting its leads to a battery of accumulators *B*, as shown in Fig. 32. The current from this battery *B* can be regulated by a resistance *R*. Let the cylinder *C* be connected to a milliammeter *mA* and a battery *V* whose voltage can be varied between, say, zero and 100 volts.

If the filament is about 10 cm. long and about 0·03 mm. in diameter, its surface will have an area of about $\frac{1}{8}$ sq. cm., and the thermionic current will be of the order of 50 milliamperes. When the potential of the cylinder is the same as that of the filament (i.e. no part of the battery *V* is inserted), no current will be recorded by the milliammeter *mA*. The current will rapidly rise, however, as more and more of *V* is inserted,

until, when V has the value of, say, 90 volts, the current reaches its value of 50 milliamperes, and no further increase will be obtained as the voltage of V is further raised. In other words, it reaches its *saturation* value when V is 90 volts.

The explanation of these phenomena is found in the effect of mutual repulsion among the electrons given off by the filament. The effect of connecting the cylinder to the battery V is to raise its potential above that of the filament and to give it a positive charge which increases as the voltage of the portion of the battery connected to the cylinder is increased. When none of V is inserted, the electrons escaping from the filament form a cloud in the space surrounding the filament, and since they are negatively charged they repel back other electrons attempting to leave the filament. A few electrons having very high velocities may have a sufficient speed to break through the cloud and reach the cylinder, but not in sufficient numbers to cause the milliammeter to deflect. If, now, the cylinder is charged positively, the electrons will be drawn away from the neighbourhood of the filament and pulled across the space to the cylinder. The effect of the charge on the cylinder is to weaken the effect of the space charge of the cloud of electrons surrounding the filament, so that more electrons can leave it, and therefore an increasing number reach the cylinder. The whole of the electrons given off from the filament do not reach the cylinder until its potential is increased to 90 volts. The form of the curve connecting the value of volts on the cylinder and the electron current reaching the cylinder is given in Fig. 33.

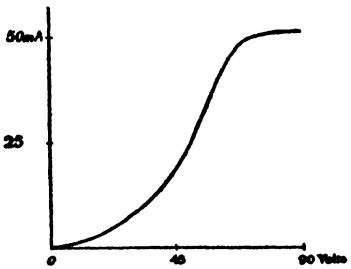


FIG. 33.

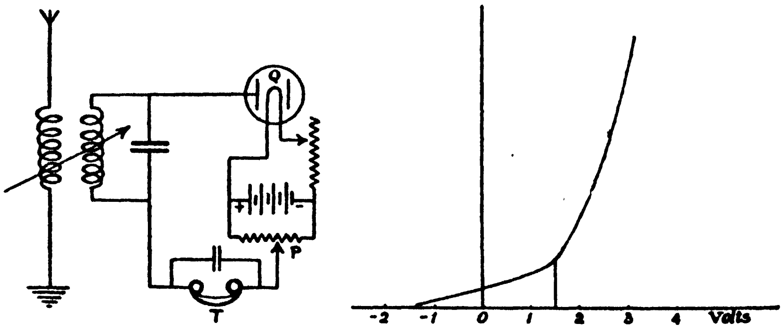
It may be observed that a similar phenomenon is found in the case of ordinary thermal evaporation from a liquid. If the evaporation is taking place in still air, the molecules which have escaped from the liquid remain near the surface and the process of evaporation is soon checked. If, on the other hand, a breeze is blowing, the molecules near the surface of the liquid are carried away and the process of evaporation is correspondingly more rapid.

Hitherto we have supposed that the filament is placed in a perfect vacuum, but if the exhaustion of the containing vessel has not been carried far enough, and there is a small quantity of gas remaining, the distribution of the space charge will be affected by another agency. Some of the electrons given off by the filament will collide with the molecules of the remaining gas; and if the exhaustion has been carried far enough the electrons travelling towards the cylinder will have attained sufficient velocity and sufficient kinetic energy to break away other electrons from the gas molecules with which they collide, and what is known as ionization by collision will take place. The effect will be (1) to increase the number of negative electrons present which are being drawn towards the cylinder; (2) to provide a number of positively charged ions (i.e. the molecules which have attained positive charges by the loss of electrons knocked out of them by collisions).

The positive ions will be driven into the cloud of negative electrons surrounding the filament, while some may be even driven into the filament itself. The general effect will be to reduce the space charge. Such ionization by collision will therefore increase very considerably the current through the valve by way of the cylinder *C*, which, since it is connected to the positive end of the battery *V*, we may call the anode.

The earliest type of thermionic receiving valve, patented by Professor J. A. Fleming in 1904, was constructed on the lines described above. It consisted of a small electric-light bulb containing a filament of tungsten or carbon bent in the shape of a hairpin. The filament was surrounded by a cylinder of nickel or copper gauze to which a wire was attached, which was brought out through a seal in the glass for connexion to outside circuits. The ends of the filament were brought out by means of wires passing through the base of the bulb, to which a battery of from 4 to 12 volts could be connected for raising the filament to incandescence. Such a valve is shown diagrammatically at *Q* in Fig. 34 (*a*), and is connected in the receiving circuits as shown. The potential between the metal cylinder of the valve and the filament can be varied by adjusting the slider of the potentiometer *P*. If we refer to the characteristic curve given in Fig. 34 (*b*) connecting the electron current through the valve and the anode potential, it will be seen that the lower portion of this curve is of the same form as that obtained in the case of a

carborundum crystal (see Chap. VII). Thus, by adjusting the valve so that it functions at the bend of the curve, the valve will act as a detector in the same way as that in which a crystal acts. This arrangement was at one time largely used by the Marconi Company as a detector, since it was found to be very robust and, unlike many crystals, was not rendered insensitive by strong signals or atmospherics. It was not, however, as sensitive as a good crystal, and until the invention



(a) FIG. 34. (b)

and development of the three-electrode valve by Dr. Lee de Forest the valves were not generally adopted.

We have seen that the current between the cylinder and filament in the arrangement in Fig. 32 varies as the space charge is varied, increasing as the negative space charge round the filament is decreased or as it becomes less negative. Thus, if a metallic electrode is introduced near the filament, between it and the cylindrical anode, any charge introduced on this third electrode or any change in its potential will modify the space charge near the filament and will produce a change in the anode current. In order that the electrons may pass through it, the third electrode must be in the form of a grid. Such a grid may either be formed by a spiral of thin wire or a fine-meshed cylinder, or it may be simply one or two zigzag wires placed parallel to the filament, these being joined in parallel if two zigzags are used. A lead from the grid is brought outside the glass bulb, so that the grid may be connected to the external circuits. The anode or plate of the three-electrode valve consists either of a metal cylinder surrounding the grid or of a flat plate.

In considering the action of the grid on the anode current of

the valve, it will be assumed that the valve is what is known as a hard valve (i.e. so highly exhausted that the effect of any gas molecules remaining in the valve can be neglected). Let us suppose, first, that the grid has a negative potential and that the plate has a large potential with respect to the filament. The effect of the negative potential of the grid will be to increase the negative space charge round the filament, and also, if the grid is a close spiral, to screen the filament to a certain extent from the action of the positive potential on the anode. Thus,

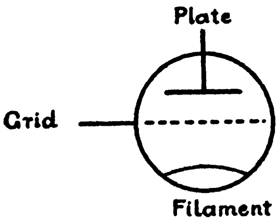


FIG. 35.

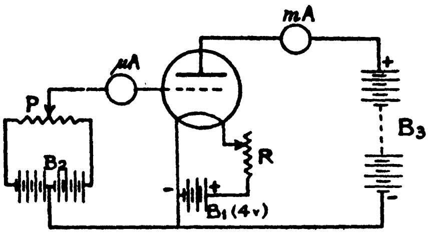


FIG. 36.

if the negative potential of the grid is sufficiently great, the electronic current to the anode may be entirely checked. As the negative potential is decreased the electron current to the plate will begin to flow, and as the grid becomes positive its effect will be to assist the anode potential in reducing the space charge and to draw more electrons to the anode. Since the grid is near the filament, a very small change in its potential may considerably increase this flow of electrons to the anode. As the potential of the grid becomes more positive, a few electrons may be attracted to the grid itself, and what is known as a grid current will flow. This, of course, will be small compared to the anode current, but in certain cases the effect of the grid current is highly important.

A three-electrode valve may be represented diagrammatically as in Fig. 35. The effect of the change of potential of the grid on the plate current of the valve, described above, can be experimentally verified by connecting the valve as shown in Fig. 36, where B_1 is the filament battery, R a regulating resistance, and B_3 the anode battery. The grid potential can be varied from a high negative value to a high positive value by the potentiometer P and the battery B_2. The anode and grid currents can be measured by the milliammeter mA and the microammeter μA respectively.

With this apparatus the characteristic curves for a valve can be traced. By giving different potentials to the grid by means of the potentiometer P and plotting these grid potentials against the reading of the milliammeter mA, the *anode-current grid-volts* characteristic may be drawn for a particular anode voltage given by the battery B_3. This is termed the 'static characteristic' curve, to distinguish it from the actual working or dynamic characteristic obtained under alternating-potential conditions. The numerical values found will, of course, depend upon the

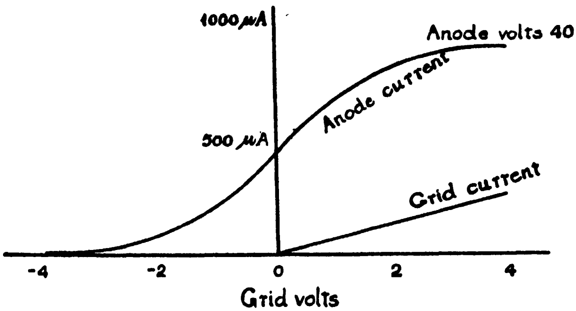


FIG. 37.

design of the valve. The general form of the curve obtained, however, is that shown in Fig. 37, the numerical values being taken only as convenient examples of the order of the quantities dealt with. If the anode voltage applied by the battery B_3 is increased, it is clear from what has already been said that saturation is obtainable with a smaller grid voltage, while the same anode current will be obtainable with a larger negative potential on the grid. Thus if the curves are traced with increasing anode voltages they take the form shown in Fig. 38. It will be seen from the diagram that the result of increasing the anode voltage is in effect to move the characteristic curve bodily to the left. Thus, below the saturation current the anode current is increased for the same grid voltage by increasing the anode volts.

Hitherto the filament current has been assumed to be constant. If, however, the filament current is increased by decreasing the resistance R, the filament will become hotter and more electrons will be given off. Thus the saturation current will be increased (see Fig. 39).

Finally, the *grid-current grid-voltage* characteristic may be traced by observing the reading of the microammeter μA in the grid circuit. This curve will be found to take the form shown in Fig. 37. It will be seen that the grid current is very much smaller than the anode current.

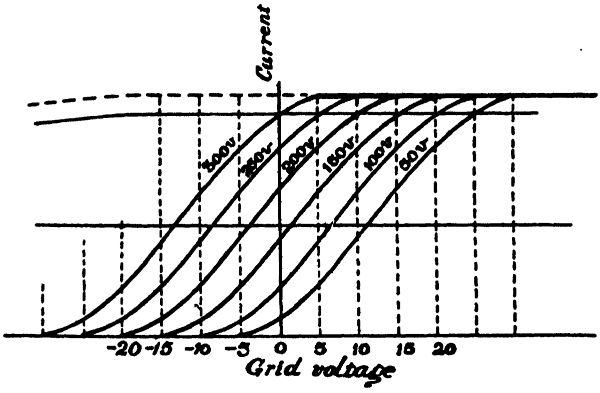
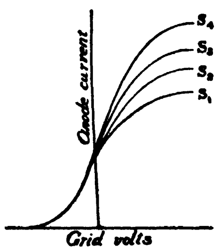


FIG. 38.



Increase of saturation current $S_1 S_2 S_3 S_4$ with increased filament currents

FIG. 39.

We have hitherto considered only hard valves, in which the vacuum is so high that practically no trace of gas remains. In the earliest valves, however, it was impossible to obtain such a vacuum, and these valves contained a trace of gas whose pressure might be of the order of 10^{-5} mm. of mercury. This gas is ionized by collision with the electrons given off by the filament. The presence of the positive ions complicates very considerably the action of the valve. They neutralize the space charge still further, and the anode current is thereby increased to a large extent. The positive ions approaching the filament may attain

a high velocity and may bombard the filament and so disintegrate it, thus shortening the life of the valve. The grid current is usually very much greater for the same positive grid potential than in the case of a hard valve, since more electrons are present. Again, when the grid potential is made slightly negative, the grid may collect positive ions on itself, thus causing the grid current to reverse. As the grid is made more negative still, it will check the electrons leaving the filament and so remove the cause of the ionization, and the grid current therefore returns to zero.

It is clear that the behaviour of a valve will change altogether with a small change in the pressure or composition of the gas present. One of the most successful soft valves was that invented by Captain H. J. Round and known by his name. This was widely used in early equipment as a sensitive detector, but great difficulty was found in all such types in keeping the gas content constant, since changes in age and in filament temperature varied the number of active ions present. To compensate for this the Round valve contained a small pellet of asbestos, enclosed in a small pocket at the top of the bulb, which on being heated gave off sufficient gas to replace that destroyed during operation. Soft valves required constant skilled attention and adjustment, and it was most inconvenient to have to restore them by heating at intervals during use. For these reasons they have become obsolete. Of recent years, however, specialized forms of gas-filled valves employing stable fillings such as mercury vapour, argon, or helium have come back into favour as power rectifiers or electronic relays. The thyratron is an example, which we shall be studying more fully later in this chapter.

Thermionic valves are now manufactured by processes of mass production. In the case of one of the best-known makes of valves, the tungsten filaments are manufactured from tungsten ores imported from Australia. The natural ore is ground into a powder, reduced by hydrochloric acid, precipitated as a low-grade oxide, and then dried. This mixture is ground with water and heated in open pans, the result being yellow tungsten oxide. This is further reduced at a temperature of 800° C. in furnaces through which a current of hydrogen passes, and a brown powder of a lower oxide is obtained. The further heating of this brown oxide at 900° C. for two hours produces a grey

metallic powder which is pressed by hydraulic presses into bars about 8 inches long and of ¼-inch-square section. These brittle bars are next heated for half an hour in a furnace and then made to coalesce by passing a current of 1,500 amperes through them in an atmosphere of hydrogen. The bars of metallic tungsten so produced are gradually beaten down with mechanical hammers to form rods about 0·03 mm. in diameter. These rods are then drawn through a series of diamond dies each smaller than the last, until after passing through the finest of them a filament wire of only 0·0005 inch in diameter is obtained. At the present time pure tungsten filaments are only employed in certain transmitting types and for laboratory use, in which cases the durability of the filament under high temperatures and anode voltages and its constancy of emission with time make pure tungsten the best material. The current-consumption of tungsten filaments has never been reduced much below 4 watts, and for receiving purposes this is capable of considerable reduction by the use of other types of filament material. It may therefore be said that pure tungsten is now only employed for certain of the larger types of valves.

An object of research has been the development of a material which will give a satisfactory thermionic emission at a lower temperature. It was shown in 1914 by Langmuir and Rogers that the addition of a small percentage of thoria has this effect. Valves with thoriated filaments work satisfactorily with the filament at a dull red heat. The current-consumption of the valve is therefore decreased, and the life of the valve lengthened. The filaments for such 'dull-emitter' valves are produced in the same way as ordinary filaments, except that the small percentage of thoria is added in the early stages.

For final exhaustion of the valves a mercury-vapour pump is used. In mass production, what is known as a 'gettering' process is used, as follows. Before the anode is finally sealed into the valve a piece of magnesium is attached to it. The filaments in the partially exhausted valves are heated to a state of high incandescence, and a potential considerably above the normal working potential is applied to the anode. The large electron current so produced makes the anode red hot and volatilizes the magnesium attached to it. The finely divided magnesium combines with the gas remaining in the bulb and deposits itself in combination with the absorbed gas on the

surface of the bulb. This deposit also absorbs any further gas liberated from the metal electrodes, &c., during the life of the valve. There are various means of volatilizing the 'getter' substance, and substances other than magnesium can be employed.

In dull-emitter valves the filament is heated for some time to a temperature higher than its normal working temperature. This has the effect of driving the thorium present in the body of the filament to the surface, where it has its greatest effect in increasing the emission current. On first heating the filament the emission is about that of tungsten, some 1 milliampere per sq. cm.; but after some hours at a temperature in the order of 2,000° C. thorium diffuses to the surface and the emission rises as much as 3,000 times, to nearly 3 amperes per sq. cm. One of the greatest difficulties experienced in the manufacture of dull-emitter valves is to prevent the thorium from evaporating in course of time from the filament, and so leaving an almost pure tungsten filament which will cause the valve to act as a bright-filament valve.

The thoriated-filament valve was a decided advance upon the plain tungsten filament in respect of the greater emission obtainable at lower working temperatures. This has the effect of increasing the life of the filament and of reducing the consumption of electrical power necessary to heat the filament. In reception, particularly where receivers may have to run for very long periods, it is most desirable that the consumption be low. There are many circumstances such as remote colonial outposts where filament current is difficult to obtain, accumulators difficult to charge, or dry batteries almost unobtainable. In broadcast reception also the running cost of a set depends largely upon the filament consumption. Current used in filament-heating is current wasted as far as the amplification of signals is concerned. The thoriated filament also assisted materially in reducing the size and cost of large power valves and the smaller transmitting valves. The ability of a valve to deliver power will be shown to depend upon filament emission as an essential factor, and hence when high power is required the filament emission must be large. This means either increased filament current or improved filament emitting efficiency.

Early research had resulted in the discovery by Wehnelt in 1904 that the oxides of certain rare-earth elements, of which thorium and barium are examples, could be employed as a

coating upon a tungsten filament; and that the emission from this heated oxide coating could be even greater than that of a thoriated wire. Patient investigation led to the development of better and better oxide coatings, and to a solution of the problem of retaining this attached to the tungsten core throughout the life of the valve. This method of cathode construction still further improved the smaller valve types, and the typical dull-emitter receiving valve of to-day most frequently employs this 'oxide-coated' type of filament. The emission from an oxide-coated filament varies from 1 ampere up to as much as 5 amperes per sq. cm. according to the method of preparation. The working temperature may lie between 850° C. and 750° C. only.

We can thus distinguish three distinct types of filament construction: the pure tungsten filament, the thoriated tungsten filament, and the oxide-coated filament. These were evolved in that order, and their emitting efficiency increases in the same order. Whereas all three types are in use to-day, there is a general tendency to adopt the latter constructions in an increasing proportion of cases. It is interesting to compare the emission and consumption of filaments at different dates in the evolution of the modern valve. The earliest types may have taken some 12 watts of energy to yield a maximum emission of 50 milliamperes. This was gradually reduced until pure-tungsten-filament valves took only 3 watts for a similar or better emission, and still lower-consumption types existed but were unsatisfactory owing to the very fragile and easily broken filament. The life of early valves varied from a matter of minutes up to a few hundred hours at most. Earlier thoriated filaments consumed perhaps 0·6 watt to provide sufficient emission for a small receiving valve, and this has been reduced until the minimum figure of about 0·1 watt has been reached. This was the case in the 2-volt 0·06-ampere filaments popular a few years ago, and valves of this minute consumption are made for special purposes where low current and portability are the chief considerations. For general reception in battery-operated sets, however, a somewhat more robust filament is found more satisfactory, and the typical modern dull-emitter receiving valve consumes 0·2 watt at 2 volts. Early valves employed almost any filament voltage that their makers found convenient, but during the War this was standarized at 4 volts

for the 'R' type of receiving valve most widely used. Later, with the advent of broadcasting, filament design progressed in two directions, one class of valve being built with a 6-volt filament in the interests of increased emission and consequently better performance, whilst another employed the 2-volt filament with a view to economy in use. The old 4-volt type fell between the two, and was eventually given up, although examples continued in use until very recent years. For a time the 6-volt valve remained the most efficient, and the 2-volt was only used when ample power to light the filament was not available. With the steady improvement in filament coatings, however, the 2-volt types gradually became equal to all others, and, owing to the convenience in using a single accumulator cell as filament supply, have now become the standard for all types of battery-operated receivers.

Up to this point we have discussed valve filaments from the historical standpoint, and have described valves heated by direct current from some form of battery. In the case of large valves, however, the filament current may reach several amperes, or even hundreds of amperes, and the filament consumption of transmitters may reach several kilowatts. These large currents are obviously unobtainable from batteries with reasonable economy. It is possible to derive them from large direct-current dynamos, and this is the source employed in many broadcast stations. It would be more convenient, however, if the filaments could be heated by alternating current, which is more readily available in many localities, and can be easily reduced in voltage by step-down transformers to any desired value. The voltage required by a filament is largely determined by the valve design, and there is not space within most valves to allow a very long filament having the necessary high resistance to operate at high voltage. The filament is usually short, for essential reasons of design; its resistance is therefore low, the current taken is relatively large, and the voltage low. Valve filaments operate at from 2 up to a maximum of some 50 volts, higher values being unusual, and large currents at low voltages can be obtained from a transformer more economically than by any other method if alternating current is permissible.

The emission from a filament depends upon the potential between it and the surrounding electrodes. For the anode current to be uniform, therefore, the filament must be at a con-

stant potential at all times, and it is an advantage also if it can be constant throughout its whole length. This is not possible in the case of direct-current heating, since the external circuits must be connected to one end of the filament or the other, and there will be a potential drop along the filament equal to the heating voltage. This is not important in the case of the low voltages used in reception, and the operating curves of the valve are specified with respect to the negative end of the filament. If, however, an alternating current is employed to heat the filament, there will be an alternating potential across it, and the potential of any point along the filament will be alternating with respect to all other parts of the circuit. Thus an alternating voltage will exist between the filament and anode, and between filament and grid, and since the anode current is controlled by these potentials there will be an alternating anode current flowing. This will vary at the frequency of the filament supply, and will introduce hum into the received radio-signals which will ruin results. It is impossible to employ an alternating-current filament supply unless some method can be found to prevent this hum and to make the filament behave as if it were at a steady potential.

One method by which this can be done is to make the point of connexion to the anode and grid circuits at the electrical centre of the filament, instead of at one end, as is usual in battery operation. Practically it is seldom convenient to employ the actual centre of the filament, which is inside the valve and inaccessible. In a few cases intended specially for ultra-short-wave use, an actual connexion has been provided and brought to a terminal outside the bulb. As it is only necessary to connect to a point which is at the same potential as the centre of the filament, this can be achieved by connexion to the centre of an external resistance joined across the ends of the filament, or to a centre tapping on the transformer winding which feeds the filament. In either case a point is chosen which is at all times exactly half-way between the potential of each end of the filament. This is often termed an 'electrical centre tapping', and is used for a variety of purposes in radio-circuit design. The feature of this centre tapping is that its potential is always exactly as much positive relative to one end of the filament as it is negative relative to the other, which will be the case no matter what the actual potential across the filament is at any

instant, or what its polarity. Thus the potential of the centre point is unaffected by changes in that across the filament, and it remains constant even if the potential across the filament be an alternating one. Another way of looking at this is that at any instant the potential of any point on the filament on one side of the centre tapping is exactly equal and opposite to that of the corresponding point on the other side; the two halves of the filament are therefore at identical potential distribution along their lengths, but opposite in sign. Any increase in the emission from one half due to its increased potential will therefore tend to be compensated by an equal decrease in emission from the other half, and the total filament emission tends to remain constant. If, therefore, the anode and grid circuits are connected to this centre point, the alternating potential across the filament will have no influence upon the average potential between grid or anode and filament, and the anode current will remain free from any alternating component which might cause hum.

This system, in which a valve filament is directly heated from alternating current, is very widely used in transmitting valves and large power amplifying valves. If the filament were truly symmetrical about the centre tapping, and no stray couplings existed between the filament circuit and those which carry the radio-signals, then no hum would be introduced through the use of A.C. Practically, however, such an exact symmetry is seldom possible, and a slight variation in the anode current at filament-supply frequency occurs. For this reason the arrangement is only used when no considerable amplification follows the valve whose filament is A.C.-heated. The attempt to use it for the earlier valve stages in an amplifier or radio-receiver results almost inevitably in slight hum, which will be amplified by the following valves until it reaches objectionable proportions. In the case of valves used to deliver large power no such amplification usually follows, and the imperfections in exact centre connexion are not sufficient to cause trouble.

From quite early days in the development of radio-reception attempts were made to operate valves entirely from the electric supply mains. As will be shown later, there is no great difficulty in obtaining anode current from the mains at sufficient steadiness, and 'battery-eliminators' have long been designed whereby A.C. can be rectified to produce direct current, and the irregu-

larities filtered out from this until it becomes steady enough for radio-work. It was in the filament-heating methods that early attempts failed. In the case of direct-current mains satisfactory results were sometimes possible. The high voltage is reduced by means of resistances, so that the current through the valve filaments is limited to the correct value, and the potential across them is therefore also correct; and provided that the mains supply is sufficiently pure the valves will operate much as if supplied from batteries. Usually, however, a proportion of A.C. is present in all D.C. supply mains, either from the original A.C. from which many are produced by rectification, or as commutator ripple from the generating dynamos; and this is sufficient to cause variation of anode current and hum. Centre tapping could be used to reduce this but is not entirely effective owing to the unsymmetrical leakages from each main to earth, which tend to prevent a perfect symmetry. Clumsy expedients were used to reduce these variations or ripples in the filament potential, such as large electrolytic condensers of several thousand microfarads across the filament terminals, 'floating' accumulators across the supply, or elaborate electrical filters; but none was sufficiently effective and practicable to put the mains-operated valve receiver upon a sound commercial basis. In the case of A.C. mains the position was even less satisfactory and the method of centre tapping inadequate to provide permanent hum-free operation of the earlier valve stages.

These failures combined with the increasing demand for entirely mains-operated receivers stimulated research, and valves were designed specially for that purpose. Success eventually came by a new method of construction in which a filament through which the heating current flows is replaced by what is termed an 'indirectly heated cathode'. In this device the two aspects of the original filament are ingeniously separated, and the electron-emitting surface is separated from the wire carrying the current which heats it. The true cathode surface from which electrons are emitted takes the form of a tiny tube coated with oxides. This forms a true equipotential surface connected to the external circuits by a separate wire joined to its own terminal and forming a single cathode connexion. The filament, now termed the 'heater', remains a tungsten wire passing through the cathode cylinder, and is in thermal contact with it whilst being electrically insulated. The insulation

between cathode and heater is made as high as possible, so that the potential of the heater has no effect whatever upon that of the cathode, and current fluctuations in the heater also have no effect upon the cathode emission. In this way it has become possible to heat the cathode from A.C. or unsmoothed D.C. from the mains without any trace of hum.

The indirectly heated valve has become the most widely used of all types, and has a number of other advantages over the older directly-heated-filament valves. In the first place the total emission from the larger surface of the indirectly heated cathode is very great. Large emission is necessary in power valves and makes for high efficiency in other types. There is a delay of several seconds before the temperature of the cathode changes with changes in the heating current, and therefore changes in emission occurring through a momentary fluctuation in the supply are eliminated. The mechanical weakness of the older filaments is eliminated, and the new valves will withstand very rough treatment or severe vibration.

It has been said that the indirectly heated cathode is a true equipotential surface. This makes for noise-free and uniform operation and removes the field, due to the filament current, which surrounded the filament of earlier valves and had the effect of reducing the emission, thus lowering the 'mutual conductance', as will be shown later. Similarly we shall see that the separation of the cathode from the heater circuits provides a very convenient method of biasing the grid. As against these advantages were a few minor defects which have now almost disappeared with increased manufacturing experience. There is a tendency for the grid to become heated by radiation from the heater, which results in the emission of electrons from the grid, and consequently in greater grid current. This is undesirable, but in general the grid current of indirectly heated valves remains higher than for the filament types. Early indirectly heated valves were also noisier in operation. This again is a defect which has disappeared except in a few isolated examples.

For reasons of convenience the heater voltage and current of these valves have been standardized in Great Britain at 4 volts and about 1 ampere, which for several years remained the only standard used in small receiving valves fed from A.C. In larger power valves the current, and in still larger types the voltage, is increased. A consumption of about 4 watts seemed the most

satisfactory for general use; but recently the increasing utility of indirectly heated valves in motor-car receivers, in aircraft and other mobile services, and in D.C.-mains receivers has led to the production of types working at from 12 to 16 volts, and at the reduced current of from 0·2 to 0·3 ampere. These can be fed from a 12-volt car ignition accumulator, or used in series with a regulating resistance on higher mains voltages. A few valves of other ratings have also made their appearance for special uses. It is interesting to note that one maker has succeeded in designing valve heaters to work at from 200 to 250 volts. These can be connected to the mains directly, taking a current of from 20 to 50 milliamperes. They have not found very wide acceptance, however, since they are unusual in some other respects and have a non-standard holder.

In America the standard heater voltage has been 2·5, and the current rather under 2 amperes. A series of valves has recently been added which take 0·33 ampere at 6·3 volts, with a wide latitude of safe variation, and these again are intended for car and mobile work. Valves of 5 and 7·5 volts also exist in American lists. During 1937 members of the Valve Manufacturers' Association decided to adopt the 6·3-volt rating as an international standard, and they now make valves to fit a new eight-pin or 'octal' holder of improved design. Thus for the first time British and American valves have become interchangeable.

Sufficient has now been said to show the changes which have taken place in the cathode-heating and efficiency of valves, and the way in which modern valves can be used equally well from direct- or alternating-current supplies. Further particulars of the mains operation of receivers will be found later when we discuss receiving circuits. It is necessary while describing recent valve types to refer to their advantages in use and the reasons which have led to their production. Where these are not entirely clear they will be found more fully dealt with later under the appropriate chapters. We shall now pass on to discuss the improvements which have taken place in the electrode systems of valves, bearing in mind that these apply equally to valves containing filaments, still used largely for battery-operated equipment, and to those which contain heaters and indirectly heated cathodes intended for mains operation.

The triode or three-electrode valve has already been described

and its typical characteristic curves given. To understand the data supplied by manufacturers illustrating the properties of a particular valve, certain other quantities connected with the valve must be introduced. One of these is the 'voltage amplification factor', μ, and is the measure of the effectiveness of the valve as an amplifier. The characteristic curves of Fig. 40 will make the meaning of this factor clear. From these we see that if the grid potential be increased by a small amount, of, say,

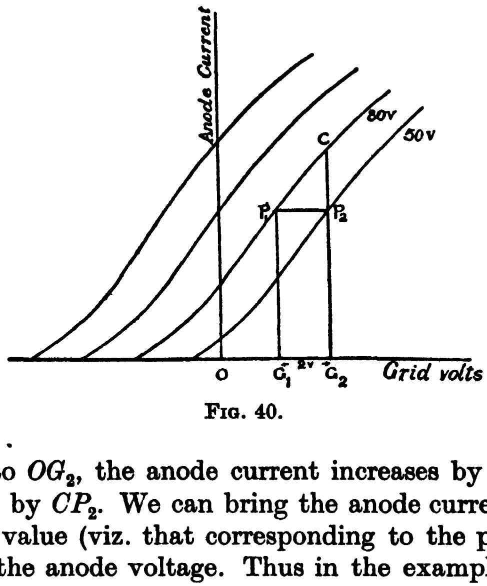


FIG. 40.

from OG_1 to OG_2, the anode current increases by an amount represented by CP_2. We can bring the anode current down to its original value (viz. that corresponding to the point P_2) by decreasing the anode voltage. Thus in the example shown in the figure the points C and P_1 lie on the characteristic corresponding to the anode voltage of 80 volts, and if the anode voltage is decreased to 50 volts the anode current corresponding to the grid voltage OG_2 will be G_2P_2, equal to the original current G_1P_1. If, therefore, by some arrangement the anode current flowing through the valve be kept constant, the change in the voltage of the anode with respect to the cathode for a change in grid voltage represented by GG or, say, 2 volts, is in our example 30 volts. The rate of change of anode voltage is 15 volts for a change of 1 volt on the grid of the particular valve considered. This, then, is the maximum change of anode voltage which could be produced for unit change of grid voltage; and for a valve with the characteristics shown the voltage amplification factor would be 15. It is clear that since the change

in anode voltage is greater than the change in grid voltage producing it, amplification has taken place; and if the portion of the characteristic over which the change occurs is a straight line, no distortion will be added to a complex wave form applied to the grid. This will merely appear in magnified form in the anode circuit.

In practice the constant anode current necessary for the above sequence of events is obtained by the insertion of a high impedance into the anode circuit. This tends to prevent any sudden change of anode current, but cannot prevent it entirely. It is not possible to maintain the anode current absolutely unchanged, and so the amplification obtainable from the valve is always less than the amplification factor. The proportion of this which can be practically utilized, and the means for doing this, will be dealt with in the chapter on amplification.

A second important quantity is the A.C. resistance of the valve, which is simply the resistance represented by the anode-cathode path (the internal portion of the anode circuit) to an A.C. current which is being amplified. It can be defined as the change in anode voltage corresponding to unit change in anode current read off from the curves at constant grid potential. It is thus a function of the anode circuit only, depending upon the construction of the valve and the anode voltage.

The third important quantity is the mutual conductance, M, which is defined as the increase in anode current for a unit change in grid voltage. This is the quantity most frequently quoted as a measure of the general 'goodness' of the valve, as it is most closely related to the practical performance. In Fig. 40 the mutual conductance is obtainable from the curves, and is the ratio CP_2/P_1P_2, or, to express it differently, the slope of the anode-current grid-potential characteristic curve. This quantity is therefore frequently referred to simply as the 'slope' of the valve. A simple and useful relationship exists between μ, the A.C. resistance, denoted by R_a, and the mutual conductance, whereby when any two are known the third can be found. For this reason valve-makers usually specify two of these factors only. This relationship is

$$M = \mu/R_a,$$

where R_a is measured in ohms, and M given in conductance, reciprocal ohms, or 'mhos'. The practical unit for M is the

'micromho'. To verify this, consider the triangle CP_2P_1 in the figure. Then if we assume that the characteristic curves shown are for increments of 1-volt grid potential in each case, we have that $M = CP_2$, $R_a = P_1P_2/CP_2$, and $k = P_1P_2$, bearing in mind that each corresponds to unity change in grid potential. Hence the relation

$$R_a = \frac{P_1P_2}{CP_2} = \frac{\mu}{M} \quad \text{or} \quad M = \mu/R_a.$$

It will now be possible to consider the three-electrode valve, commonly termed a triode, in relation to its development and modern characteristics. We have seen that amplification factor and A.C. resistance are related, so that for a valve of given mutual conductance one cannot be increased without the other. This expresses partially the practical fact that in valve manufacture it is most difficult to increase the one without decreasing the other. But since it will be shown later that the mutual conductance is roughly a measure of the utility of the valve as an amplifier, then there must be a constant struggle to improve valve design in the direction of the lowest possible ratio of R_a to μ. Early triodes of the tungsten-filament type had an amplification factor of from 5 to 20, and an A.C. resistance of the order of from 10,000 to 50,000 ohms, the mutual conductance being well below unity. It was of course possible to increase μ by increasing the control of the grid over the electron stream, which could be done by closer grid mesh and less distance between grid and filament; but in so doing the A.C. resistance was also raised to from 100,000 ohms to over 1 megohm. This gave little real advantage, however, since to make use of the increased amplification factor it is necessary to employ at least an equally high resistance in the external anode circuit, which we shall see later is inconvenient if not actually impossible. Thus the actual amplification obtainable from early triodes was poor.

Years of steady development in electrode design, coupled with the reduced A.C. resistance yielded by the indirectly heated cathode, have improved this considerably. At present we have indirectly heated triode types of which the following are typical: the low-amplification or 'L' type, having a μ of about 16 and R_a 6,000 ohms; a medium type or 'HL', having a μ of 35 and R_a 12,000 ohms; and a high-amplification or 'H'

type of μ about 90 and R_a 35,000 ohms. Until quite recently such valves held the field of radio-reception very completely, but they have been increasingly replaced by improved types having more than three electrodes. The remaining field in which the triode still has a prominent place is that of power-output valves designed to deliver several watts of power to the loud speaker, and still larger types employed as power oscillators and amplifiers in transmission. In the latter work the triode is still supreme.

The power which a valve can deliver bears a proportion to the maximum which can be safely dissipated in heat at the anode, and is therefore often expressed in terms of safe anode dissipation. This may lie between 6 and 60 watts in the case of receiving output valves, and very much higher values in the case of the large amplifiers used in public-address and cinema equipment, and in transmission. The largest triodes differ very much from their early ancestors and from receiving types. They are frequently made either of silica and metal to withstand great heat, or from metal throughout. The latter are demountable, and can be taken apart when necessary for repair or cleaning. They are permanently attached to evacuating plant whereby the vacuum can be maintained against possible slight leakage, whilst the anodes are water-cooled to allow of larger dissipation. A variety of expedients are adopted in smaller valves to increase the safe heat dissipation at the anode. Amongst these is the use of carbon as an anode material, or the division of the anode into a number of smaller metal strips having an increased total radiating surface. Air-blowers may be used to cool the glass envelope of such smaller valves, and are practicable up to several kilowatts, whilst water cooling is preferable at larger powers.

In reception the triode has certain defects, the most serious of which is the presence of electrical capacity between the anode and grid, which are two metal masses in close proximity. This capacity forms a coupling between the grid and anode circuits able to transfer energy back from the anode to the grid, and it will be shown that this is the normal method of setting valve circuits into continuous oscillation. In an amplifier this oscillation is a defect and limits the stable amplification obtainable from the valve. When triodes were in general use for reception, somewhat complex circuit arrangements were

employed to neutralize the effect of this capacity and to stabilize the valve, but these were inconvenient and amplification was still limited. Also it was desired to raise the ratio of μ to M in order that better amplification might be obtained at the higher frequencies.

Research directed along these lines led to the evolution of the screen-grid valve. This was the first type to depart radically from the original idea of a simple anode, cathode, and control grid, and to introduce additional electrodes; although valves having two similar grids had been previously introduced for other purposes but without lasting success. The screen-grid valve, however, proved of the greatest importance and resulted in a minor revolution in receiving technique.

The idea of screening the grid electrostatically from the anode was intended primarily to reduce the capacity between these two electrodes. The screen consists of a grid-like structure which separates the anode from the grid and cathode more or less completely. From the point of view of alternating potentials the screen is 'earthed' to the cathode through the low reactance of a large condenser, and therefore whilst both grid and anode have a comparatively large capacity to the screen, they retain very little capacity to each other. The screen is maintained at constant potential and cannot transfer potential changes from anode to grid. It will be interesting to form an estimate of the reduction in anode-grid capacity which results from the insertion of an earthed screen into a normal triode receiving valve. The anode-grid capacity of a triode might be of the order of 15 mmfd. The addition of a suitable screen to this valve would result in an anode-screen capacity of perhaps 20 mmfd., whilst the grid-screen capacity might be 6 mmfd. Since the screen can be treated as if at cathode potential, it may be said that the capacity of the electrodes to cathode has been somewhat increased by the presence of the screen, a minor drawback when working at very high frequencies. As against this, however, the anode-grid capacity will be reduced to the order of 0·002 mmfd., or even less if the screen be very well designed. This is a reduction of several thousand-fold, and extremely valuable. The figures given here are of course approximate, since the effectiveness of the screening varies considerably with different valves, but they represent the kind of improvement to be expected.

As a result of this enormous reduction in anode-grid capacity the screen-grid valve is much more nearly a one-way device than is the triode. Whilst changes of grid potential still control the anode current, there is a very much smaller inverse effect, and the alternating potential transferred back to the grid as a result of variations of anode potential is small. When the valve is in use for the amplification of comparatively low frequencies, between the limits of zero and about one megacycle per second, this capacitive coupling can generally be neglected; whilst at still higher frequencies it may not be entirely negligible, but may not be sufficient to set up self-oscillation of the valve and its circuits. Thus the design of radio-receivers has been greatly simplified because it is no longer necessary to provide special circuits to 'neutralize' the internal coupling of the valve, and also because greater amplification can be obtained from a single stage without instability being encountered. It is of course essential to prevent coupling between the grid and anode circuits through causes external to the valve if this increased amplification is to be made use of, and elaborate steps are often necessary to do this.

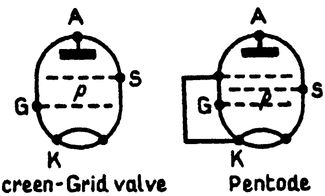


FIG. 41.

We must now consider the screen-grid valve from a different point of view, that of its characteristic performance curves and voltage relationships. Although the screen was originally conceived for the reasons just mentioned, it has a beneficial influence upon the efficiency of the valve as an amplifier. In Fig. 41 is shown the diagrammatic representation of a screen-grid valve, from which its construction is also broadly illustrated. The inner grid, which corresponds to that of a triode and is similarly proportioned, is frequently termed the 'control grid' or 'grid No. 1'; here we shall term it simply 'the grid' except where greater clearness seems necessary. The outer grid is really the screen structure, usually grid-like and more or less completely surrounding the inner grid and cathode assembly. It will probably be of quite close mesh. This screening grid is often called 'grid No. 2', but for the present we shall term it 'the screen'. If a negative potential were applied to the screen, or if it were merely joined to the cathode from the direct-potential

as well as the alternating-potential point of view, the 'control' of this grid over the electron stream would be so great that little or no anode current could flow. A positive steady potential is therefore always applied to the screen when the valve is in use, and, as has been mentioned, steps are taken, such as by a large condenser between screen and cathode, to ensure that this screen potential remains constant whatever may happen to the grid or anode.

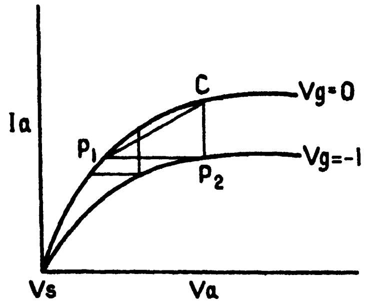


FIG. 42.

Now if a positive potential V_s be applied to the screen and a small negative potential V_g to the grid, these two electrodes will resemble an ordinary triode valve exactly, and would have a similar characteristic if the screen were treated as an anode. If now a higher positive potential of about double V_s be applied to the true anode, say V_a, then electrons will be attracted through the mesh of the screen to form an anode current, and if this mesh be suitably fine it can be regarded as screening the space p in the diagram (between it and the grid) from the effect of the anode voltage. The anode can thus only attract electrons from those which would otherwise be attracted to the screen, and anode current flows only at the expense of screen current. Hence the sum of the anode and screen current tends to be a constant equal to the current which would flow to the screen if the outer anode were not there and the screen behaving as the anode of a simple triode valve. This condition enables us to form an idea of the way in which the anode current of a screen-grid valve should vary with increasing anode potential. For V_a equal to or less than V_s we should expect no anode current to flow, since all electrons emitted from the cathode will be drawn to the screen. As V_a is increased above V_s anode current will commence to flow and will increase quickly, but as V_a becomes considerably higher the anode current will tend to rise more slowly and will reach a steady maximum when all the electrons reaching the screen are attracted through it to the anode. The curve we should expect would resemble that shown in Fig. 42.

Suppose this to represent the curve when a steady control grid bias of 0 volts be applied, and the second curve that for a bias of -1 volt. The curves have a very similar slope, and if we construct the triangle CP_1P_2 of Fig. 40 it will be seen to differ very much in shape from that of the triode, being far more acute-angled towards the apex at P_1. Now, it has been shown that CP_1 measures the mutual conductance of the valve, which is seen to be quite good and of the same order as for the triode. P_1P_2 measures the amplification of the valve, which is clearly very large, whilst the ratio of P_1P_2/CP_1 which measures the A.C. resistance is also considerable.

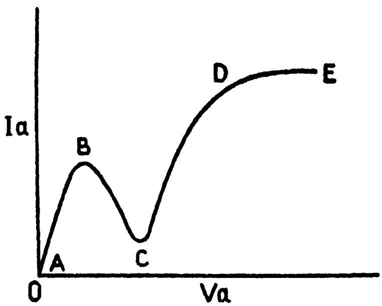


FIG. 43.

The screen-grid has thus quite different characteristics from the correspondingly constructed triode, and in many ways superior ones. The amplification factor is very high indeed, lying between 1,000 and 2,000 for practical examples. This extreme figure cannot be entirely utilized because the A.C. resistance is also considerably higher, lying in practice between about 200,000 ohms and over 1 megohm. The mutual conductance is quite high, however, this showing that a large proportion of the amplification factor may be expected to provide useful amplification, and it is a fact that screen-grid valves considerably exceed the triode in practice. A gain of from 200 to 300 is quite possible from a single valve working under favourable conditions.

The characteristic curves of actual screen-grid valves differ somewhat from that which we have just estimated, showing a curious kink, in the manner of Fig. 43. The peak at B is due to an effect termed 'secondary emission' from the anode. The first few electrons to strike the anode as its potential is raised do so with such a high velocity that they knock off fresh or 'secondary' electrons from the material of the anode. Temporarily the anode acquires the characteristics of a partially emitting cathode. Between the points B and C on the curve there is insufficient anode potential to retain these secondary electrons, which are attracted back to the screen and merely

augment the screen current at the expense of the anode current. This is the cause of the dip in the curve at *C*. Beyond *C* the anode potential is sufficient to retain secondary electrons in its immediate vicinity and finally to attract them to itself; therefore the curve from *C* onwards through *D* to *E* is of the shape expected for a screen-grid valve.

The irregular shape of the practical curve at low anode voltages is of no importance in normal amplification, since by operating the valve at a sufficiently high anode potential and with negative grid bias the lower kinks in the curve are entirely avoided. It is interesting to note that over the portion of the curve *BC* the anode current falls for increasing anode potential. This is the condition of apparent negative resistance discussed later in relation to the arc transmitter; and if an oscillatory circuit be connected to the anode of a valve working over this region, continuous oscillation at the frequency of the circuit will occur provided that damping is not excessive. A valve used in this manner is termed a 'dynatron' oscillator and, whilst not widely employed for generation of oscillatory power, has a number of practical applications. The screen-grid valve works well in this manner as a rule when the anode potential is below that of the screen, a condition to be studiously avoided in more normal operation. Since the effect depends upon secondary emission, however, and this is a very variable factor from valve to valve, it is found that some specimens favour dynatron oscillation very much more than others, and may show wide variations during the life of the valve.

After the development of the screen-grid valve, investigation naturally turned to the problem of further improvements along similar lines, and to the possibilities of adding still further electrodes to the simple triode. Valves of this kind are termed pentodes, heptodes, octodes, and so on, the first syllable of the name indicating from a Greek or Latin root the total number of electrodes within the valve. It is usual to number the successive grids 1, 2, 3, 4, in the order of their distance from the cathode, that nearest to this being grid No. 1. Normally grid 1 will be the control grid, and some other a screen. We shall not describe all these multi-grid valves here in full, since their function is so closely bound up with their uses. They have been evolved to meet specialized needs in modern receiving circuits, and can be best described when these circuits are themselves described.

One of these valves, termed the pentode, merits special attention, however, as it has become a type of primary importance. The kinks in the screen-grid valve characteristic have been said to make little difference to its use as an amplifier. If, however, an attempt is made to draw power from the anode circuit of a screen-grid valve, as for the operation of a large loud speaker, it becomes necessary to swing the anode potential through wide limits. The power existing in the anode circuit when an alternating potential is applied to the grid will vary as the product of the anode-potential variations and anode-current variations, and for these to be large it is necessary to employ a greater length of the curve, and to encroach upon the irregular region. We must find a way to suppress the effects of secondary emission and to straighten out the lower portion of the curve. This has been done in the pentode by the addition of a third grid between the anode and screen. This grid is connected to the cathode directly and is of fairly open mesh so that it has little effect upon the normal action of the valve. It serves, however, to trap secondary electrons which might otherwise pass from anode to screen. These strike the new grid, termed the 'suppressor grid', to form a grid current which returns harmlessly to the cathode. The suppressor grid can be regarded as a device which removes the kinks from the screen-grid curve, and thus allows the anode potential to be varied over wide limits without encountering negative resistance or other bad effects.

The pentode retains a large measure of the high amplification factor of the original screen-grid valve whilst being capable of considerable power output from the anode circuit, and is a very favourite output valve in modern receivers. It has a somewhat different behaviour from a triode when delivering power to a load, which will be considered in detail later, but for the moment we will note that there is a tendency for the pentode to deliver equal changes of anode current at all frequencies and over a wide range of load conditions. This has a favourable effect on the power obtainable, and in general a pentode will deliver more oscillatory power for a given anode voltage than a triode. There is a tendency, however, for a somewhat greater degree of distortion to occur.

The alternating-current resistance of pentodes is somewhat lower than that of the simple screen-grid valve, and it has recently been found possible to obtain better amplification from

them in the stages of modern receivers devoted to high-frequency amplification. A variety of special pentode types having characteristics appropriate to various circuit conditions are becoming increasingly used. Recently it has also been found possible to develop pentodes which will safely withstand high anode voltages and currents and are thus able to deliver considerable power to the anode circuit load. These valves have an application both in the production of very large low-frequency signal power suitable to the driving of large loud-speaker installations, and also to radio-frequency power amplification, this making them an important addition to the range of modern medium-power transmitting valves.

In recent years a wide variety of multiple valves has been evolved. These consist of two or more sets of normal electrodes arranged around a common cathode of ample emission and enclosed in a single glass container having a sufficient number of terminal connexions or pins to provide individual contact to all these electrodes. Such valves can be regarded as simply a number of normal and independent valves combined for convenience into a single unit, and offer only the slightest technical variations from a group of separate valves. They have been produced for commercial reasons which include economy in construction, simplification in the wiring, and reduction in the bulk of receivers; and they represent a commercial rather than a technical advance. Multiple valves normally comprise two or three sets of electrodes composing valve units likely to be closely associated in the circuit of a typical receiver, and their titles should show quite clearly what they comprise. We shall refer later to the most frequent combinations when discussing receiver design. Typical examples are the double-diode triode, which as its name implies contains in one bulb two independent diodes and one triode valve; the diode or double-diode pentode; and the triode hexode.

It will be noticed that diodes occur several times in the above list, and frequently in pairs. This may occasion some surprise when it is remembered that the diode was the original thermionic valve and was rendered almost entirely obsolescent by the improvements which followed the introduction of the grid. It is true that for a number of years the diode vanished almost entirely from radio-reception, and was seldom met with except as a power rectifier, since the superior sensitivity of the grid

detector seemed to place it in the forefront of progress. With the development of multi-grid valves, however, it became easy to produce great amplification in the other stages of a receiver, and the detector became important not so much for its sensitivity as for the perfection of its rectifying action. With the development of broadcasting, high-quality reproduction entered for the first time into radio-reception, and it was found that grid detection introduced an objectional degree of distortion. The long-neglected diode proved to be the most distortionless rectifier available, and, since its lack of any inherent amplification was no longer a drawback, returned quickly to favour. The properties of diode rectification will be considered later, but it may be mentioned now that these include very complete rectification, minimum distortion, large output, the ability to handle large input potentials without overload, and lack of discrimination between varying degrees of modulation depth.

There is no necessity for a diode detector to provide a heavy anode current or to possess large cathode emission, and it is clearly wasteful to employ triode valves to do the work of diodes. Hence special small diode valves have made their appearance, adapted to the rectification of radio-signals. Since many modern circuits employ two diodes, and the cost and bulk of these is so small, it has become general to make up a double-diode valve in which two small anodes are provided adjacent to a common cathode. Taking advantage of the small emission needed it has also been found practicable to include diode elements within the envelope of more elaborate valves. Thus two small anodes may be included near an unused portion of the cathode of a pentode valve, constituting one of the double-diode-pentode combined valves previously mentioned. As such combinations are quite satisfactory, the single diode is not very often used.

There is an entirely different field in which the diode has always remained supreme, that of power rectification. This was unimportant for some years, but with the increasing use of alternating-current mains-operated receivers has come a group of large diode rectifiers specially adapted to the rectification of mains current for receiver high tension. These valves are made in both single- and double-diode types, and since these are suitable for half- and full-wave rectification respectively they are termed half-wave and full-wave rectifiers. In common with

all other valve types under discussion, these rectifiers are constructed with both directly heated filaments and indirectly heated cathodes, and since hum is not so important in a power rectifier the filament types are widely used, and can be heated from alternating current. Power rectifiers of this type exist in the widest range of sizes. Those used in reception range from a small valve delivering up to some 60 milliamperes at 250 volts to those delivering 120 milliamperes or more at at least 500 volts. Larger sizes almost without limit are used in large amplifier installations and in transmission.

The rectifiers just described are of the high-vacuum type. In their electrode design will be found a large cathode of ample emission and anodes capable of dissipating safely heat caused by the current flowing through the internal resistance of the valve. The anode-cathode spacing will be made sufficient to prevent flash-over at the highest potentials which can occur across the valve in use. Large valves may employ carbon as the anode material and silica or metal as the envelope, whilst additional anode-cooling by water- or oil-circulation is sometimes necessary. Those valves employed in receiving equipment, however, resemble large forms of the simple diode. Special types exist for the rectification of very high potentials at moderate currents, such as may be needed in television receiving equipment, and these are not necessarily large valves but are characterized by large electrode-spacing.

It is now frequently preferred to fill the valve with mercury vapour in place of the original high vacuum. Such rectifiers are termed mercury-vapour rectifiers, and have the ability to handle larger currents in proportion to their size. In addition the internal resistance of the valve is very much lower, as conductivity is increased through the ionization of the mercury vapour. This is an important factor since it implies a very much smaller voltage drop within the valve and less variation in the output voltage with changes of output current. Current and voltage tend to become independent, as when using a secondary battery, which is necessary for the ideal operation of certain circuits. The mercury-vapour rectifier is therefore gaining ground and appearing in more and improved types.

Whilst considering the use of mercury vapour within the valve, mention should be made of a recent development of importance, the 'thyratron' or gas-filled triode valve. Its

properties differ essentially from the vacuum triode, being more akin to those of an electronic relay. If anode potential be applied to a thyratron whilst the control grid is biased negatively, no anode current flows. As the grid is made increasingly positive, however, a point is reached at which anode current commences to flow, when ionization immediately sets in. This ionization is rapid and cumulative, resulting in the anode current's rising to its maximum value in a very brief period. Once this ionization current has been set up, however, the control grid loses all effect upon it, and it can only be stopped by the momentary removal of the anode potential. The action is thus quite different from that of the triode valve, since whilst a small change of grid potential can set the anode current flowing, a reduction to the original or a still lower grid potential cannot reduce the anode current to its former value. The device thus resembles a non-self-resetting relay. Its extreme rapidity of action, approaching one microsecond, together with the comparatively large anode currents possible, has made the thyratron a very valuable device. Its applications, however, lie largely outside the field of radio-communication, although it may be found in associated circuits such as those for remote control. Mercury vapour is not essential to the thyratron and may be replaced by any readily ionized gas such as helium or argon. These gases are used when it is desired to speed up the action of the relay to the greatest possible extent, without necessarily such a large anode current, and have applications in high-definition television equipment.

At the present time there is under development a new device akin to the valve and which may prove superior to it in certain applications, notably that of amplification. This is the electron-multiplier evolved in an attempt to simplify the amplification technique of television transmission, but which seems to possess properties that may make it invaluable in many fields. It is claimed that amplifications of one million are possible from a single multiplier tube, whilst at the same time uniform amplification is possible over a very wide band of frequencies. Thus the tube might replace a six-stage valve amplifier whilst yielding improved reproduction. The associated circuits and internal design are complex, however, and for these reasons it is not possible to estimate how far the multiplier may replace the ordinary valve in radio-communication.

The electron multiplier makes use of the secondary emission effect mentioned as a defect in the screen-grid valve, and is probably the first device to do this successfully. A very simple construction of multiplier is shown diagrammatically in Fig. 44, which illustrates the principle upon which all work. A few electrons are emitted from a normal filament or cathode K and strike a first anode A_1. This anode is maintained at a suitable positive potential so that an electron current will flow to it as in a simple diode; but the anode is not constructed of the usual materials, having a surface treated to yield maximum instead of minimum secondary emission. Surfaces of the type used in photo-electric cell construction have this property, one of the most efficient being caesium oxide deposited upon a silver base. It is possible with the latest technique to prepare surfaces which will emit as many as ten secondary electrons upon the impact of a single 'primary' electron travelling with sufficient velocity.

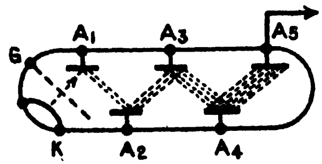


FIG. 44.

The action of the cathode and first anode of the electron-multiplier may therefore be regarded as similar to that of an ordinary valve, and several methods are possible whereby the electron stream striking the first anode can be modulated by the signal to be amplified. One method which would do this is illustrated in the sketch, a control grid being inserted between the cathode and anode exactly as in a triode valve. Other methods are employed in practice, such as variation of the first anode potential, the effect of varying externally produced electrostatic or magnetic fields, or the production of the initial emission by the impact of light upon a photosensitive cathode. However this modulated emission is obtained, it reaches the first anode and causes the emission of secondary electrons. These are several times greater in number than the primary electrons striking the first anode, and when these secondary electrons reach the second anode A_2 an amplified anode current will flow. The second anode is maintained at a higher positive potential than the first, and thus attracts the electron stream to its surface; but this second anode is also treated to promote secondary emission, and so the secondary electrons from the first anode behave as primary electrons to the second anode,

giving rise to a still greater number of new secondary (or tertiary) electrons. The multiplier is so constructed that the electron stream strikes a number of successive anodes, five of which are shown in the sketch. Larger numbers can be employed in practical multipliers, up to at least ten, and each of these successive anodes is maintained at a higher positive potential than its predecessors. The final anode will therefore be at a very high positive potential relative to the initial cathode, and this is a minor practical drawback of the tube. In passing from anode to anode the electron stream collects a very large proportion of secondary electrons, and the anode current will be larger at each successive anode. At the final anode it is drawn off to the output circuit. The emission of secondary electrons is at all times strictly proportional to the number of primary electrons giving rise to it, and the final anode current is a true copy of that at the first anode, but greatly amplified.

A simple construction such as that illustrated would not be very efficient as there is no provision to guide the electron stream from each anode to the next. Since the tendency is for the secondary electrons to be emitted in all directions indiscriminately, steps must be taken to guide the electron stream so that the bulk of it reaches each electrode. In the construction shown many electrons would miss the anodes entirely. The leading workers developing the multiplier, who include Zworikin of the Radio Corporation of America, Farnsworth, also an American, and engineers of the British Baird Television Company, have introduced elaborate arrangements of subsidiary electrodes for this purpose. These electrodes are generally negatively charged and so repel the electron stream from places where it should not pass. Their field tends to compress the stream inwards upon its own axis, and is thus said to 'focus' it into a beam. The process thus broadly outlined is termed 'electron-focusing' and will be referred to again as we study the cathode-ray tube, in which it plays a vital part. Space will not allow a description of more advanced multiplier tubes. Research is still in an early stage and it is impossible to say at present how far it will be possible to develop them.

One further device remains to be described under the general title of thermionic valves. This is the cathode-ray tube, long of value to the laboratory worker and now becoming of great practical importance in the reception of television. Whilst not

strictly a valve, the tube has so much in common with valves both in operation and construction that it can be best regarded as a specialized type. The purpose of the cathode-ray tube is to make an electron stream perceptible, not as an amplified electric current which can produce sound in a loud speaker, but as visible light. The modulation of this stream then becomes apparent as changes in illumination rather than as acoustical or mechanical energy.

In common with the valve, the cathode-ray tube employs a heated cathode of very similar nature. This is frequently in-

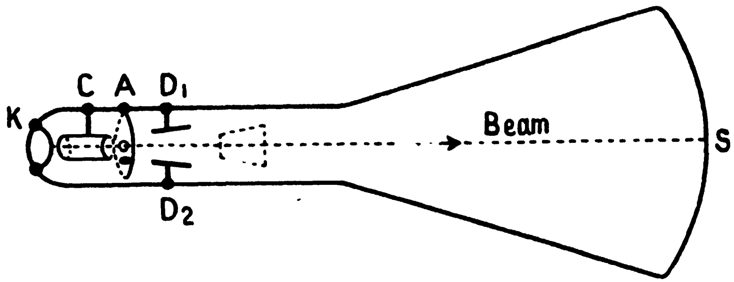


FIG. 45.

directly heated by alternating current. It is preferable that emission occur from a single point, and to assist in that direction early cathodes consisted of a thick tungsten strip filament upon which was deposited a spot of oxide coating, so that emission was largely confined to this spot. Cathodes so constructed had a working life of comparatively few hours, and the modern tendency is to employ cathodes similar to those in indirectly heated valves, but smaller. From the cathode, marked *K* in Fig. 45, a stream of electrons is emitted towards the positively charged anode *A*, which in its simplest form consists of a ring or plate pierced with a central hole through which a large proportion of the electron stream passes. Some provision is necessary to concentrate the electron emission from the cathode so that a large proportion of it reaches the centre of the ring anode, as without this much would be wasted on the walls of the tube. The negative or Wehnelt cylinder effects this. It comprises a metal cylinder *C* forming a tunnel between the cathode and anode, and, although not shown in the diagram, frequently enclosing the cathode almost completely. It is charged negatively relative to the anode (and usually to the cathode also)

by a potential which may lie anywhere between zero and a few hundred volts, according to the design of the electrode system. The field of this negative cylinder repels the negatively charged electron stream, thus concentrating it along the axis of the cylinder and preventing side emission. Most of the electrons pass towards the anode, are considerably accelerated by its positive potential, and pass through the central hole at high velocity to form the 'cathode ray'. Thus we see that the so-called 'ray' is merely a narrow stream of rapidly moving electrons, and whilst it behaves in many ways like a beam of light or electric waves it must on no account be confused with these. The cathode ray differs from the electron stream in a normal valve only in that the electrons composing it are travelling at high velocity in parallel paths to form a narrow stream, not unlike a ray, whilst in the valve they travel more or less broadcast. The essential construction of cathode, negative cylinder, and anode is characteristic of cathode-ray tube design, but it may be found difficult to recognize these components in tubes of some makes. A wide variety of electrode designs exist with a view to improved performance and better concentration, whilst some makers introduce additional electrodes or very considerably vary the shape of those mentioned. The essential principle, however, consists simply of a negative field to concentrate the emission from a cathode so that the bulk of it passes through an anode space which is at high positive potential and therefore accelerates the stream of electrons into an effective beam. The whole of this electrode assembly producing the beam is sometimes referred to as an electron 'gun'.

Whilst describing the formation of this it will be a convenient opportunity to digress for a moment to mention that at the present time beam emission is being applied to normal valve design with striking results. Until very recently no attempt was made to concentrate the electronic emission so that it passed completely through the electrode system, but now valves have been tried in which an electron beam is built up in the manner just explained and then passes through grids to anode in the usual manner. This concentration of the electron stream results in smaller and more efficient electrode designs, and enables the beam to be far more effectively controlled. 'Beam valves' are as yet in a semi-developed state, but the first commercial type to be released upon the American market is a power output

valve termed the '6L6'. Two of these in class-B push-pull are capable of a signal output power of 60 watts with very small input, and this is some four times that possible from the previous types under similar conditions. In Great Britain the leading valve-designers are known to be working on similar lines and have produced striking degrees of amplification in the laboratory. Few commercial types have yet appeared, but they may be expected to represent a major advance in valve efficiency. The adoption of beam construction is leading to a form of pentode of improved characteristics but needing no suppressor grid to overcome secondary emission, because the latter is negligible in intensity relative to the beam.

The cathode ray, having been produced, passes down the body of the tube as shown to strike the fluorescent screen *S*. This is a specially prepared chemical coating inside the enlarged end of the tube, and makes use of the properties of certain chemical substances to fluoresce under the impact of high-velocity electronic bombardment. A small spot of light is formed upon the screen at the point where the beam strikes it, and in this way the beam becomes visible. The brightness of the spot will vary with the velocity and number of the electrons, and it can move over the screen with great rapidity to show any movements of the beam itself. For a spot of given area the brightness and colour depend upon the nature and preparation of the screen material, will vary in direct proportion to the number of electrons comprising the beam, and will increase with beam velocity, but in the latter case not according to a linear law. The original screen material was willemite, or zinc orthosilicate, which fluoresces very brightly a deep-green colour. Other materials have since come into use, some of the most important being zinc sulphide, which gives a blue-green glow, cadmium tungstate, giving a blue, and zinc phosphate, which yields a deep-red glow. The materials are chosen to give a coloration of screen suited to the work for which the tube is required, the greens or reds being suited to visual observation whilst the blues are more actinic for photographic purposes. Modern screens seldom employ a pure substance but rely upon mixtures, from which it is possible to obtain a nearly white fluorescence, considered desirable for television purposes.

A second factor often more important than colour is the time lag of the screen material, illustrated in Fig. 46. From this curve

it will be seen that maximum fluorescence is reached in from 1/1000 to less than 1/10,000 second from the impact of the electron beam and can thus be considered as instantaneous for the great majority of applications. Fluorescence does not vanish immediately after the removal of the beam, however, but continues in the form of an exponontially diminishing afterglow for a time which varies widely for different materials. The time taken to fall to 0·64 of the initial brightness is taken as the time lag or time constant for the material, and an average value is

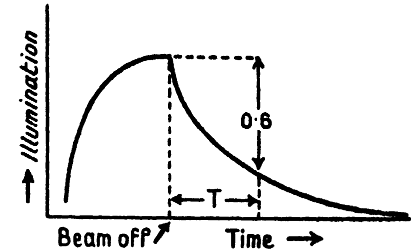


FIG. 46.

that given of 1/25 second. Materials exist which have an afterglow of only a few microseconds, zinc tungstate being an example, whilst on the other hand the red-glowing zinc phosphate has the longest afterglow known, of some 30 seconds. When the tube is in use for visual observation of brief or transient phenomena this long afterglow is most valuable, since it enables a phenomenon lasting a few microseconds or less to be examined for half a minute after it has transpired. In television, on the other hand, very rapid response is necessary, and screen materials of minimum afterglow must be chosen.

We have seen how the cathode ray is produced and how it excites light upon the fluorescent screen. In all applications of the tube the glowing spot is employed as a tracing pencil which can trace out on the screen curves showing the variation of one quantity with another. To do this motion must be given to the beam in strict proportion to the quantities which are to be examined. Fortunately two simple methods exist for deflecting the beam, the electric and the magnetic field. The beam is simply a flow of electrons and thus has all the properties of an electric current flowing along a conductor. It will therefore experience a force when passing through an electric or a magnetic field, and this force will be proportional to the product

of the current and field strength just as in the case of a simple wire. We can arrange that the field be applied so that the resulting force acts perpendicularly to the direction of the beam, and under this condition it can be shown that the electrons travel through the field in a curved path and that the resulting change of direction of the whole beam is proportional to the applied field, provided of course that the electron velocity remains constant. The beam can therefore be deflected by two electromagnets situated outside the tube, and if the fields of these be mutually perpendicular the beam will trace out upon the screen the graph showing how the current in one magnet is varying in relation to the other. If these currents be related to other physical properties, or if one of them be caused to change linearly with time, it will be clear that an unlimited range of relationships can be investigated. Used in this way the tube is an invaluable tool to the laboratory worker and replaces the earlier mechanical forms of oscillograph, with the advantage of an almost infinite rapidity of response.

Very frequently the beam is deflected by electrostatic fields produced between two metal 'deflector plates', marked D_1 and D_2 in Fig. 45. These may be simple rectangles of metal spaced sufficiently apart to allow free passage of the beam at its extreme limits of deflexion, or may be curved to follow the path of the beam. The latter results in a somewhat closer plate spacing, more intense field, and better deflexional sensitivity. The sensitivity of a tube in this respect may be expressed in volts/cm., referring to the displacement of the spot upon the screen for a potential applied between the two deflecting plates at a stated anode potential. Where electrostatic deflexion is employed in both planes a second set of plates is provided in the plane perpendicular to the first, being indicated by the dotted outline. These may be termed D_3 and D_4. All plates will of course be brought out to terminals either at the base or sides of the tube, the latter being preferable when working with high frequencies owing to the reduced lead capacities obtainable. It is quite possible to combine electrostatic deflexion in one plane with electromagnetic in the other, and this arrangement is often convenient.

Consider an electron moving with velocity v through an electrostatic field of strength E, such that the lines of force are perpendicular to its path. If the charge on the electron be

e, it will experience a force given by Ee perpendicular to the line of motion. Let the mass of the electron be m. Then its acceleration in the direction of the field can be found from the relation Force $= ma = Ee$, whence $a = Ee/m$.

In unit time the electron will have been deflected out of its original straight path by a perpendicular distance given by:

$$\text{Displacement } s = \tfrac{1}{2}at^2 = \tfrac{1}{2}Ee/m.$$

This is equivalent to motion in a curved path, as shown in

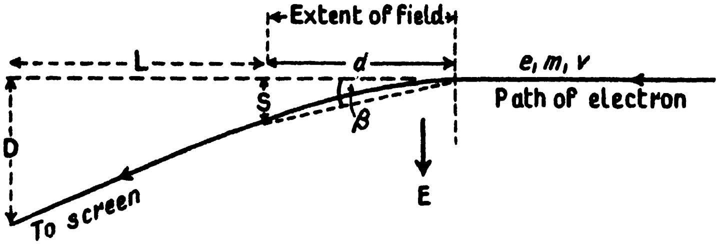


FIG. 47.

Fig. 47, the radius of which can be shown to be given by $r = mv^2/Ee$; but if the deflexion be taken as very small and the unit time interval also small, its direction of motion will have been changed through an angle β approximately expressed by:

$$\tan\beta = \beta\ (\text{approx.}) = \frac{\tfrac{1}{2}Ee/m}{v} = \frac{Ee}{2mv}.$$

If the electron now passes out of the field and continues to travel in a straight line for a distance L before reaching the fluorescent screen, it will have been deflected by a distance D on the screen given by:

$$D = L\tan\beta = E(e/2mv)L.$$

Now e and m are constants for the electrons which make up a cathode ray, L is a constant for any particular tube, and v constant under given conditions of anode voltages, &c. Thus it is clear that the deflexion of the beam shown upon the screen is directly proportional to the deflecting field strength E, fulfilling the necessary condition for an undistorted tracing upon the screen of varying potentials applied to the deflector plates. Errors can occur if the field E is not uniform over the whole area through which the beam is deflected, and it is necessary

that the electrode system be symmetrically constructed and suitably proportioned to give this condition.

If the deflexion cannot be taken as small, but the electron stream travels through a uniform field for a distance d in an appreciably curved path, it can be shown that expression for D becomes:

$$D = E(ed/mv^2)(d/2+L).$$

Here again the deflexion remains proportional to the deflecting field strength E, and therefore to the potential between the two deflector plates.

In the case of magnetic deflexion, let the magnetic field strength be H, proportional to the current flowing through a deflecting electromagnet. Then the force acting upon an electron perpendicularly to its original straight path becomes Hev. The radius of curved path can be shown to be mv^2/Hev, and the final expression for the deflexion becomes:

$$D = H(ed/mv)(d/2+L).$$

An important point which becomes evident from these expressions is that for a small deflexion by either system the movement of the beam depends upon its velocity, which is in turn mainly determined by the anode voltage employed. The deflexion for a given field strength will be inversely proportional to anode voltage. Thus when the latter is increased to provide a brilliant screen fluorescence, the sensitivity of the tube to deflecting potentials is reduced. This fact is taken into account when describing the deflexional sensitivity of particular tubes. It is not sufficient to specify a screen displacement of a certain number of millimetres per volt p.d. between the plates, unless this were given for a range of anode voltages. Instead, the ratio of deflexion to anode potential is stated. Thus if a tube is described as having a sensitivity of $500/V$ mm. per volt, this means that if the maximum anode voltage in use were, for example, 2,500 volts, then a deflecting potential of one volt would produce a movement of the spot on the screen of 500/2,500 or 1/5 mm.; 50 volts would thus be necessary to deflect the beam 1 cm., and if the width of the screen were 20 cm. we should need 1,000 volts to sweep the beam completely across it. This is in fact the order of deflecting potential needed by the tubes used in television reception to-day.

The simple electrode systems described would not be suffi-

cient to produce a sharp tracing spot upon the screen unless additional means were provided for 'focusing' the beam. The chief methods employed are known as 'gas focusing' and 'electron focusing' respectively. The former has the merit of simplicity, whilst the latter is better suited to high beam speeds, and is considered essential for high-definition television.

A gas-focused tube contains essentially the electrode system shown in Fig. 45, but in place of a high vacuum the envelope contains traces of inert gases which are ionized by the passage of the electron stream. As a result of collision between high-speed electrons and a few heavy gas molecules, free electrons and a number of relatively massive positively charged ions are produced. The former are repelled by the electrons composing the beam, and together with these find their way eventually to the walls of the tube. A conducting path is desirable whereby they may return to the cathode as the 'beam return current', and the inside of the glass bulb may be metallized or coated with colloidal graphite to facilitate this process. In certain tubes the screen has been deposited upon a conducting surface fitted with an external circuit to cathode.

The positive ions, on the other hand, tend to remain in the path of the beam, where they form a positive core that can attract electrons in towards the axis. They thus have the effect of keeping the beam narrow, and prevent electrons from straying out into the surrounding space. Excellent focus is obtainable from this effect, but since ionization depends upon beam velocity and the number of electrons present, critical adjustment of anode voltage and cathode emission is necessary to arrive at a condition giving a sharp spot. This may not be serious for laboratory work, in which the gas-focused tube can be very useful at moderately low frequencies. The increased beam current brought about by ionization enables satisfactory operation at quite low anode voltages, a bright trace being possible at as little as 400 volts, when the tube is also sensitive and the equipment simple. In television, however, it is not desirable for beam current to affect focus, since any attempt at modulation will lead to a loss of sharpness.

A second more serious defect is that of time lag. Ionization is not an instantaneous process, and a brief interval is needed in which the electron stream can ionize sufficient gas molecules to build up an adequate focusing field. Should the beam be

deflected at so great a rate that it does not remain in any one position long enough for ionization to occur, focusing will not be complete. Gas focusing therefore becomes poorer at high rates of movement, and is unsatisfactory when the beam is required to traverse the whole screen more than about one million times per second. At high radio-frequencies it is not possible to examine the wave form with conventional types of gas-focused tube, and the spot ceases to be sharp well below the speeds required for high-definition television scanning. It is interesting to note, however, that recent research has shown possibilities of great improvement in this direction, and that by the use of new types of gas filling and quite different pressures tubes have been made which respond adequately at television frequencies. Once again it is not possible to state definitely what future may lie in store for the gas-focused tube, or to affirm that it is necessarily obsolete.

The second process of electron-focusing is based entirely upon the effects of electric fields upon the beam. The highest possible vacuum is necessary within the electron-focused tube, and since there is no question of gas ionization or any other process involving time delay the beam remains focused at the highest rates of motion. In this respect the tubes are ideal and well suited to high-frequency measurements and television; but, just as so few technical improvements are entirely without drawbacks, electronic focusing of high precision is not a simple process and tends to increase the cost of the tubes and the complication of external circuits. Large tubes focused in this way necessitate very considerable anode voltages, of from two to twenty thousand. The compression of scattered emission into an approximate beam by the negative cylinder is a simple example of electronic focusing, and in this type of tube will be found a system of additional electrodes producing an analogous effect. In general there will be two or three successive anodes at progressively high positive potentials, the adjustment of which controls the focusing action. There may also be one or more negatively charged electrodes in addition to the simple negative cylinder. The operation of this electrode system is governed by a new branch of science termed 'electron optics', dealing with the behaviour of electron beams under complex electrostatic and magnetic fields; for focusing is equally possible through the effects of magnetic fields produced outside the

tube for convenience. The name indicates a fact which has emerged from recent research, that the electron beam follows very similar laws to those of optics. If the beam be regarded as replacing a ray of light, then the curvature of the lines of force composing a field through which it passes is equivalent to a lens of similar curvature, and the intensity of the fields is

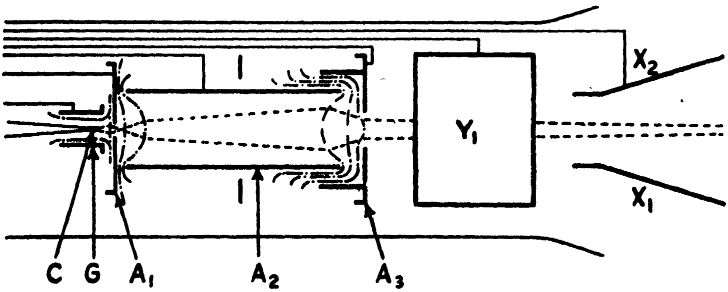


FIG. 48.

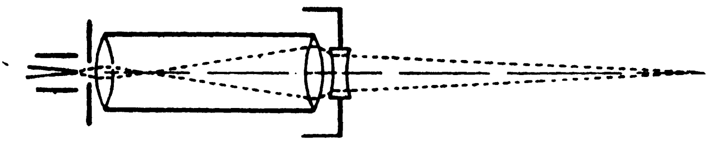


FIG. 49.

equivalent to refractive index; then the mathematical relationships which hold are of the same general form. As an example Fig. 48 shows a sketch of the lines of force surrounding the negative cylinder and positive anodes on the assumption that this is surrounded by a uniform weaker field. The electrode system shown will be further explained in Chapter XVI. Fig. 49 illustrates the optical lens system to which the electrostatic field system of Fig. 48 can be compared. It will be seen that a divergent beam of electrons entering the field is rendered convergent in a very similar way to a beam of light passing through a thick biconvex lens. This fact, that 'electronic lenses' correspond to thick optical lenses, and moreover that the electric field will be continuously varying throughout the area concerned and thus resemble a lens composed of material having variable refractive index, implies that the equivalent optical calculations will be highly complex. In practice it is seldom possible to apply electron optics rigidly, and the optical parallel serves merely as

a valuable guide to the type of electrode construction likely to prove effective. Details of design are worked out by trial-and-error methods.

The new field of electron optics is receiving intensive study and has led to big improvements in cathode-ray tube design. As an example of its utility, an electron microscope has been designed and is slowly reaching perfection, in which an electron-optical system equivalent to the microscope is evolved from the focusing action of electrical fields. It is well known that the resolving power of a microscope increases inversely as the wavelength of the light with which the objective is illuminated. The ultra-microscope takes advantage of this fact to see objects too small for observation by visible light, through the agency of ultra-violet light. Similarly, the use of an electron stream gives the beneficial effects of a still shorter equivalent wave-length, comparable to the size of an electron, and thus may be expected to make visible very much smaller objects than is possible with the best optical microscopes. Electron optics provides means to focus the electron stream in electron-multiplier design, in the new beam valves, and in the electron cameras used in the latest television equipment.

Modulation of the intensity of a cathode ray was not needed in early work, deflexion of the beam only being required. With the use of the tube for television, however, it becomes essential to modulate the beam without loss of precise focus, so that the light and shade of an image can be reproduced without loss of definition. This is not an easy matter, and the most general method has been to impose a modulating bias upon the negative cylinder. In the case of gas-focused tubes this mainly affects focus. The adjustment is such that the beam remains in exact focus when at maximum brightness and so the brightly illuminated portions of the image remain sharp, whilst darker portions are produced not only by a reduced beam intensity but also by considerable defocusing. The latter spreads the beam, reducing the brightness of the fluorescence, but causing a loss of sharpness in the darker areas. Modulation by negative cylinder potential is also used with electron-focused tubes, in which case the effect upon focus is much less and the result superior. Here modulation controls the number of electrons which are concentrated by the cylinder through the anode aperture, and hence the screen illumination. It also controls the direction at which the

bulk of these pass into the anode region, and modifies the beam velocity. These factors are critical in gas focusing, but in the electronic case are less so since the major focusing process occurs after the beam has passed the first anode, and is little affected by anything which happens to it before that point is reached. Neither of these modulating processes is perfect, and much work has been done to produce better systems, an idea of which will be given in the later chapters on television.

EXAMINATION QUESTIONS

1. Draw typical examples of the following characteristic curves:

(*a*) The anode-voltage anode-current curves of a diode for different filament currents.

(*b*) The anode-current grid-voltage curves of a triode for different values of anode voltage.

City and Guilds of London Institute. Preliminary Exam. 1934.

2. Describe the electronic and ionic actions which occur inside a triode. What is meant by space charge, and how does it vary between cathode and anode?

Institute of Wireless Technology. May 1935.

A valve is stated to have an impedance of 3,000 ohms and an amplification factor of 15. What information does this convey to you? What is the mutual conductance of this valve? If the anode current is 25 mA. with an anode voltage of 200, calculate the anode current for an anode voltage of 150 and the same grid bias.

I. W. T. November 1934.

4. Explain the principal stages by which the filament of a valve has been evolved, from the simple tungsten wire to the modern indirectly heated cathode. What are the reasons for these changes? Give the main advantages of an indirectly heated cathode, and mention cases in which it would be undesirable to employ this construction.

5. Why is the addition of a screen-grid an improvement to the triode amplifying valve? For what purpose is the suppressor grid added to convert this into a pentode, and what other methods have been adopted in certain cases whereby similar advantages are obtained without this additional grid?

6. Write a short report of about 150 words on the electron-emitting materials commonly found in valves. *I. W. T. June 1937.*

7. Explain what is meant by the amplification factor of a three-electrode valve. Draw a diagram of a circuit that can be used for the measurement of this factor. (Where must the grid of a triode with cylindrical electrodes be placed so that the amplification factor is a maximum ?) *A. I. W. T. June 1937.*

8. Draw a diagram showing typical characteristic curves for a three-electrode valve, and indicate how the values of the three main parameters of the valve may be obtained from it. *A. I. W. T. June 1937.*

9. Describe the two-electrode thermionic valve, and explain its uses in wireless telephony. *Grad. I. E. E. 1936.*

10. What reasons have led to the introduction of more than one grid into thermionic valves ? Describe the characteristics and uses of valves having two and three grids. *Grad. I. E. E. 1936.*

11. What is a pentode ? Explain the functions performed by the various electrodes. Give a typical anode-current grid-voltage characteristic for a pentode. *C. and G. Intermediate 1934.*

12. Describe with a sketch the construction and properties of (one) of the following multi-electrode valves, and explain with a diagram how it is used in a receiving circuit:

(*a*) A pentagrid converter.
(*b*) A double-diode triode.
(*c*) A pentode triode. *C. and G. of L. I. Intermediate 1936.*

CHAPTER VI

TRANSMISSION OF CONTINUOUS WAVES

HITHERTO we have considered only the production of damped waves: we must now turn to the consideration of the methods of producing undamped or continuous waves.

The maximum voltage which can be allowed to be developed in an aerial is a voltage a little below that which would cause the aerial insulation to break down or 'brush' discharging to take place. If we refer to Fig. 2, p. 5, we see that with a train of damped waves the maximum voltage allowable is only developed at the first oscillation of each of the groups, i.e. at points I_1, I_2, I_3, &c. The aerial, therefore, is only radiating energy at the maximum permissible rate for a small portion of the time during which the transmitting key is pressed. A great advantage is gained if the amplitudes of the oscillations in the aerial are maintained constant; for by this means a greater amount of power will be radiated for a given maximum voltage, and, conversely, for a given amount of power to be radiated from an aerial the maximum voltage which is necessary will be less. Greater efficiency is therefore possible with continuous waves than with damped waves. Secondly, the waves radiated are purer in form and, with careful arrangements in the transmitting apparatus, one frequency only is radiated when sending telegraphic signals.

The selective effects of tuned circuits of low resistance can also be used to the full in the receiving apparatus, and not only can the desired signals be received most efficiently, but interfering signals from other stations can be cut out to a very much greater extent than is possible with damped waves. It is mainly because of this reduced interference that continuous-wave working has gradually become universal. Were this not so it would be impossible to accommodate the large number of radio-stations which now exist, and serious interference between stations would be inevitable. A continuous carrier wave is also essential for telephony and television. We will now review briefly the earlier methods by which continuous waves were produced. The electric arc was first applied to produce oscillations by Duddell in 1900, but he failed to obtain oscillations of

radio-frequency. These were not obtained until 1903, when Poulsen not only succeeded in producing oscillations of very high frequency, but also in introducing considerable power into them. The complete explanation of the action of the arc when applied for this purpose is complex, and in some particulars is somewhat uncertain.

If two electrodes, one made of copper and the other of carbon, are connected to a source of high direct-current voltage and the carbon electrode is allowed to touch the copper electrode, a large current will flow across their point of contact, and each electrode will become very hot. Indeed, so much heat is generated that the copper will melt unless it is hollow and arrangements are made for it to be cooled by a continuous flow of water through it. When the carbon electrode which is attached to the negative terminal of the direct-current supply becomes hot, a large number of negatively charged particles will be driven off from it. If the carbon electrode is slowly withdrawn from the copper electrode, the negative particles will be driven by the applied voltage across the gap towards the copper electrode. These negative particles will collide on their way with intervening molecules, thereby ionizing the air and producing other charged particles. In this way a gaseous arc is formed, the current through it being carried by the negative and positive ions produced. The arc can be lengthened by drawing out the carbon electrode until the energy supplied to it from the direct-current source is sufficient to make good the energy radiated away as heat. On further lengthening the arc ceases.

Duddell found that if a circuit consisting of a coil and condenser was connected across the arc, electrical oscillations were set up in this circuit. The production of these oscillations is due to the fact that the arc has what is called a *negative resistance*, which means that as the current through the arc increases, the potential drop across it becomes less instead of greater; as it would do if Ohm's Law applied.

This phenomenon of apparent negative resistance is of great importance, and we shall see that it is a necessary condition for the production of electrical oscillations. When a circuit exhibits negative resistance, no matter how obtained, it is capable of producing oscillations at the natural frequency determined by the inductance and capacity of the circuit, or of any more suitable circuit connected to it.

It has been shown that the damping of any resonant circuit becomes less as the positive resistance and other losses are reduced, and the amplitude of any oscillations occurring therein tends to increase, whilst the time which they take to die away when once started also increases. If the resistance and losses of the circuit could be reduced to zero, any oscillation once set up in it would continue at unchanging intensity for ever, since there would be no loss of energy taking place which could damp it out. Now, it is not possible in practice to reduce the losses of a circuit to zero, but it is possible to reach the effect of zero resistance if energy is fed into the circuit from outside in a sufficient quantity to make up for all that energy dissipated in heat or radiation. If still more energy than this be supplied from an outside source, the resistance and circuit losses will be more than compensated for, and there will be a gradual accumulation of energy in the circuit. This corresponds to a condition of negative resistance, and the form which this extra energy takes will be either that of heat or of electrical oscillations if the circuit be capable of these.

Negative resistance implies that an increase in current round the circuit corresponds to an increase in potential also. Thus the product of current and potential tends to rise indefinitely, instead of being a finite product as in normal circuits. This product is a measure of energy, and hence the energy of a circuit having negative resistance tends to increase indefinitely. Of course this energy must come from somewhere, and in the case of an oscillating arc it comes from the battery or dynamo which supplies the arc with current. Under practical conditions the energy lost from a circuit increases with that present in it, and the greater the oscillatory energy in a resonant circuit the greater will be the losses from radiation and all other causes. Thus for a circuit of given negative resistance energy tends to build up until the total losses equal the total rate at which energy is supplied. The circuit will then continue in oscillation with constant amplitude for an indefinite period, and continuous waves will be radiated from it.

If we introduce into a direct-current circuit two large inductances or choke coils, the current flowing in the circuit tends to be kept constant, since the inductance of the chokes will present a large resistance to any changes in the value of the current. When an oscillating circuit containing an inductance L and a

condenser C is connected across the arc, as in Fig. 50, the voltage across the arc will cause a current to flow into the condenser. Now since the current in the direct-current circuit is constant, the current flowing into the condenser must come from the arc, and therefore the current across the arc itself must decrease. This fall of current will be accompanied by an increase of the potential across the arc, and hence the voltage of the condenser will rise. The increase of the condenser voltage will continue until it is equal to that across the arc, but since the current through the coil L is continually changing, the effect of the self-induction of the coil will be to produce an E.M.F. tending to cause the current to persist. This E.M.F. will charge the condenser still further until a potential maximum is reached, when the condenser begins to discharge. The current from the condenser circuit will now increase the arc current, and therefore decrease the voltage across the arc. The result will be that the condenser will be completely discharged and, as we have seen already in considering oscillating circuits, a charge of the opposite sign to the original charge will be built up on its plates until a maximum voltage of the opposite sign is obtained. Oscillations will therefore take place, which will persist as long as the arc is kept burning.

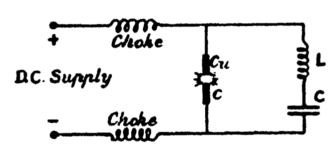


Fig. 50.

When high-frequency oscillations are required, it is necessary that the voltage changes should follow very closely the changes of arc current. To ensure this, special precautions must be taken to keep the arc as cool as possible and to enable the heat produced to be rapidly radiated away.

The arrangements introduced by Poulsen for this purpose were: (1) burning the arc in a powerful transverse magnetic field; (2) burning the arc in an atmosphere of hydrogen or hydrocarbon gas; (3) water-cooling the copper anode and the sides of the arc chamber.

With the arc system a special method of telegraphic signalling had to be used. It is not possible to make and break the supply circuit as with the spark system, since the arc would go out each time the circuit was broken. The usual method employed is so to arrange the sending key that, when closed, it short-circuits a portion of the coil known as the 'spacing coil' introduced in

the resonant circuit. By this means the closing of the transmitting key changes the self-induction of the coil and alters the wavelength to which the circuit is tuned. In this way two waves slightly different in wavelength are sent out continuously. One wave, called the 'spacing wave', is sent out when the sending key is open, and another, called the 'marking wave', when the key is closed. At the receiving station the apparatus is adjusted to receive the 'marking wave' only. This method of signalling remained in use for many years when any large oscillatory current was to be keyed, and is still occasionally met with.

Concurrently with the development of the arc system of transmission considerable attention had been paid abroad to the development of high-frequency alternators. These are machines designed to produce currents of radio-frequencies which can be applied directly to the aerial, just as ordinary low-frequency alternators at a power station are used to supply current to the mains. Ordinary alternators produce currents of frequencies from perhaps 25 to 100 cycles per second, while alternators for wireless purposes must produce currents of at least 15,000 to 20,000 cycles per second.

In a low-frequency alternator used for ordinary power or lighting circuits, currents are generated by rotating a coil of wire in a magnetic field, or in larger machines by the reverse process of causing a rotating magnetic field to cut across fixed conductors. Obviously, to obtain the vastly greater frequencies required for radio-transmissions many difficulties of design have to be overcome. The chief of these are: (1) the enormous speeds at which the moving parts of the machine must run; and (2) the great losses which occur from the induction of stray eddy currents in the metal work of the machines due to the rapidly changing magnetic fields.

Both these difficulties are best met in a type of machine known as the inductor type of alternator, and the most successful high-frequency alternators have been developed on this principle.

Fig. 51 represents a portion of an ordinary alternator in which a steady current from the fixed windings produces north and south poles, *N* and *S*, whose magnetic field induces an E.M.F. in the armature windings which run perpendicular to the paper and are indicated by $\ominus$ and $\oplus$, according to whether the currents produced in the conductors of the windings are

towards or away from the reader. The alternator shown in Fig. 51 (*a*) is transformed into an inductor alternator by transferring the windings from the moving armature to the slots in the fixed stator, as shown in Fig. 51 (*b*). The strength of the field cutting the armature windings then varies according to whether a slot or a tooth of the rotator, which is made of steel, is opposite the conductors of the winding. The rotator itself

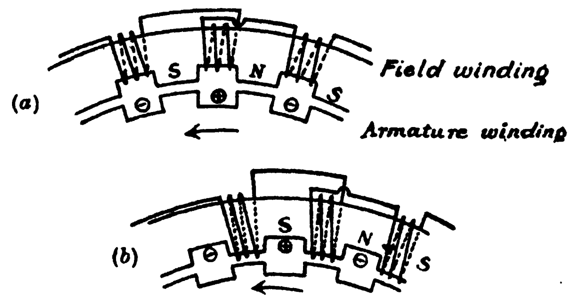


FIG. 51.

carries no winding, and the mechanical difficulties of running it at high speeds are therefore reduced.

The Alexanderson alternators, which were made up to sizes of 200 Kw. for frequencies of from 20,000 to 100,000 per second, were of this type. The Alexanderson alternators at the New Brunswick station had a rated output of 200 Kw. generated at a frequency of 22,100 cycles per second when running at 2,170 revolutions per minute. There were 64 armature windings in all in two halves of the machine, each of which generated 35 amperes at 130 volts. The whole 64 windings were collected in the same number of separate primaries of a large air-cored transformer, the single secondary of which delivered the whole output to the aerial. The voltage at the terminals of the secondary of the transformer was about 2,000. In the running of high-frequency alternators it is of the greatest importance that the speed should be kept constant. A variation of ½ per cent. in the speed of the 200-Kw. machine would cut the aerial current down to half.

Alternators of somewhat different design from that of the Alexanderson machine have been developed in France by Latour and Bethenod. A difficulty of the Alexanderson design is that, in order to ensure very high frequency, the pole pieces

have to be arranged very closely round the stator, so that the gaps between the poles become very small and the difficulty of arranging the winding of the armature increases. In the French machines this difficulty was overcome by using three alternators on the same shaft, so that successive poles are found on different alternators, and the space left by the missing poles on each alternator allows of the coils of the armature being better placed. By taking the magnetic flux on the side of the pole teeth into account, it is possible to recombine the three alternators in one machine. The theory of the process is, however, somewhat complicated.

A special feature of the French machines was that they were enclosed in air-tight cases and run in a vacuum in order to reduce windage losses. Two 500-Kw. alternators of the Latour-Bethenod type were installed in the French station at St. Assise.

In early types of the Goldschmidt alternator high-frequency currents were obtained by employing step-up stationary frequency-changers to increase the frequencies of current from the original alternator. This type of alternator, however, was hardly used outside Germany.

The development of the thermionic valve has provided more convenient means of generating continuous high-frequency currents for transmitters of all powers than either the arc or alternator, and has now entirely replaced both in the equipment of new stations, whilst those which employed the older methods either have been or will soon be changed over to valve generators. The Poulsen arc and Alexanderson alternator played their important parts in bringing the present system of international radio-communication into existence, and have been described because of their theoretical and historical interest; but the remainder of this chapter will be devoted to the valve oscillator. It is the latter only which has made possible the economic generation of short waves, and thus so widely extended the range of radio-transmission.

We have seen that, if oscillations are introduced into a circuit tuned to the correct frequency, currents and potentials are built up in the circuit, whose magnitude depends on the resistance and losses which have to be overcome. If we could design a circuit without resistance or losses, then not only would large oscillations be built up, but the oscillations would persist longer, since there would be no damping effect to reduce their energy. Now, with

the thermionic valve it is possible to supply energy from the anode circuit to the grid circuit in order to make good the energy of the oscillations excited, which is otherwise wasted in overcoming the ohmic resistance and other damping losses of the circuits. This process is termed 'reaction', or in earlier writings retroaction or regeneration.

One method of doing this is by mutual induction between a coil introduced in the anode circuit and the coil of the grid circuit L_1C_1, as illustrated in Fig. 52, where L and L_1 are

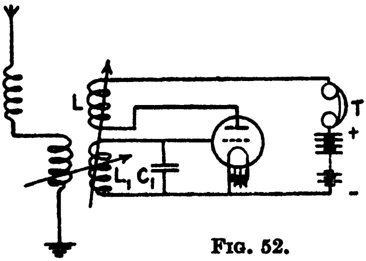


FIG. 52.

coupled together. The high-frequency currents in the grid circuit cause changes of potential to be applied to the grid. These changes of potential are nearly 90° out of phase with the currents producing them (see Fig. 53 (*a*) and (*b*)). The oscillations of grid potential cause in turn oscillations of anode current which are in phase with the grid potentials (see Fig. 53 (*c*)). If the inductance of the anode circuit is small the anode current lags 90° behind the current in the coil of the grid circuit. Thus the current through the reaction coil L in the anode circuit induces in the grid circuit an E.M.F. which is in phase or 180° out of phase with the current in the grid circuit, according to the manner in which the coil L is connected, i.e. according to whether the mutual inductance between L and L_1 is positive or negative (Fig. 53 (*d*)). If, therefore, the coil L is connected the right way round, energy will be fed into the circuit L_1C_1 which will maintain the oscillations in that circuit.

If R is the resistance of the grid circuit, the energy consumed in the resistance is RI^2. If the frequency of the oscillations is $f = \omega/2\pi$, and L_1 is the self-induction of the coil, then the potential v_g applied to the grid is given by $\omega L_1 I$. Assuming that the valve is working on the straight part of the characteristic, the anode current i_a will be proportional to the grid potential, and

$i_a = av_g$, where a depends on the internal design of the valve. Hence $i_a = a\omega L_1 I$.

The E.M.F. induced in the grid circuit by the mutual induction M of the two coupling coils is therefore $\omega M \times a\omega L_1 I$ or $\omega^2 aML_1 I$. This E.M.F. is in phase with the current flowing in

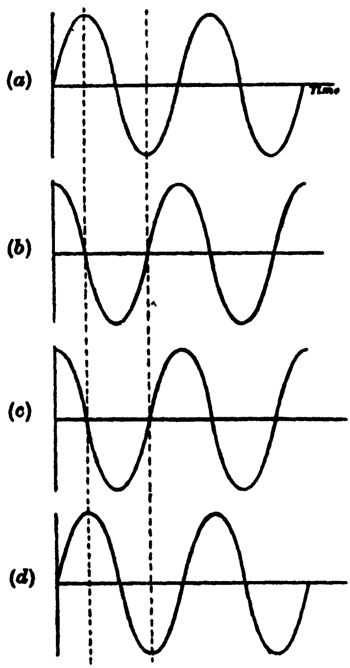


FIG. 53.

the grid coil, and therefore energy is contributed by the mutual induction equal to $\omega^2 aML_1 I^2$.

Thus the energy losses of the grid circuit are reduced from RI^2 to $(R - \omega^2 aML_1)I^2$ or to $(R - aM/C)$ since $\omega^2 LC = 1$.

By the reduction of the resistance losses of the circuit the decay of the oscillations through damping is decreased.

The reduction of the resistance R must, of course, make the resonance curve sharper. Hence the use of reaction provides a means of making a receiver more selective and hence a means of cutting out an interfering station.

It will be clear from the above that the resistance losses are further reduced as M increases, and that if the mutual induction is made large enough the second term of the above

expression may become greater than the first and the effective resistance of the circuit may become negative, which we have seen means that instead of the oscillations in the grid circuit dying away they will continue after the initial disturbance which gave rise to them has ceased. Thus any small impulse may start small oscillations in the circuit which will build up into larger oscillations. In these circumstances the valve becomes a generator of oscillations, and these phenomena form the basis of the use of the thermionic valve as a transmitter.

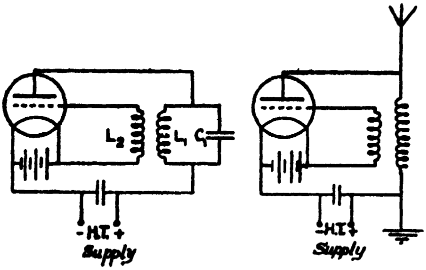
There are other methods by which reaction effects can be obtained, besides the method of inductive coupling described above. The energy of the anode circuit can, for example, be fed back through a coupling condenser between the circuits. In particular, the small internal capacity formed by the grid and plate of the valve itself may produce important reaction effects, and also the stray capacities between the connecting leads. In receiving circuits reaction is a state of things which may need to be avoided, but it is the foundation of the application of the thermionic valve to radio-telegraphic or -telephonic transmitters, particularly of the simpler types.

Oscillations are maintained in the circuit L_1C_1 in the anode of the valve (see Fig. 54 (*a*)) by the action of coil L_2 in the grid circuit, provided that the E.M.F. produced by mutual induction between L_1 and L_2 acts in the right direction and that the coupling is close enough. If an aerial and earth, or any other form of load, be added in parallel with the capacity C_1, we obtain a very simple form of transmitter, as shown in Fig. 54 (*b*). This arrangement, however, has the disadvantage that the filament battery is at a potential different from that of the ground by the value of the high-tension battery. If, on the other hand, the filament battery is earthed and a condenser is inserted in the earth lead of the aerial, the mean potential of the whole aerial is raised to that of the high-tension battery. This arrangement is equally inconvenient.

The difficulty can be got over by connecting the high-tension battery through a choke coil to the anode of the valve, as shown in Fig. 55. A blocking condenser C_1, large enough to allow an easy passage for the high-frequency currents, is introduced between the anode and the aerial circuit, so that the high-tension battery is not short-circuited through the aerial coil. This circuit is said to be 'parallel fed', whilst those of Fig. 54

would be termed 'series fed'. C_1 is sometimes called the anode stopping condenser.

In this arrangement, if the current through the choke coil varies, an E.M.F. due to the self-induction of the coil will be produced, which will tend to maintain the current at its steady value. For example, if the current through the valve is decreased suddenly, an E.M.F. will be developed tending to oppose the drop in the current, and the potential of the anode of the valve will be increased thereby. Similarly, if the current tends



(a) FIG. 54. (b)

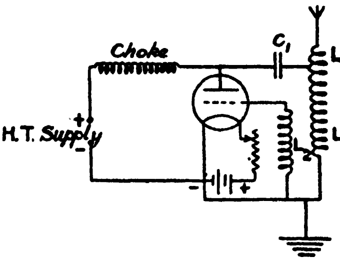


FIG. 55.

to increase, the potential of the anode will correspondingly fall. Hence the effect of the anode choke coil is to cause variations of anode potential due to variations of valve currents to be applied to the aerial or load circuit, and to keep them out of the high-tension supply.

From the valve characteristics already considered it is evident that when the grid filament voltage becomes more positive the valve current increases, and when it becomes more negative the valve current decreases. Thus when the grid potential is made more positive the anode potential is decreased, and when the grid potential is made more negative the anode potential is increased. The grid potential and the plate potential are, in fact, practically 180° out of phase.

If we suppose that by some means the circuit is given a shock, such as by switching on the filament current or closing some other switch in the circuits, then, provided the aerial resistance is low, oscillations will be produced in the aerial. These oscillations will rapidly die away through the damping of the aerial circuit, unless energy is supplied by the valve to maintain them.

It follows, however, from what has been said above, that impulses of E.M.F. will be applied to the aerial circuit by the

action of the choke coil, which will be correctly timed to build up and maintain the oscillations, provided that the coils L_1 and L_2 are so arranged that their mutual induction acts in the right direction. The growing currents in the aerial will repeatedly react through the coupling coil, and the aerial current will be built up to a maximum value corresponding to the saturation value given by the *anode-current grid-voltage* characteristic of

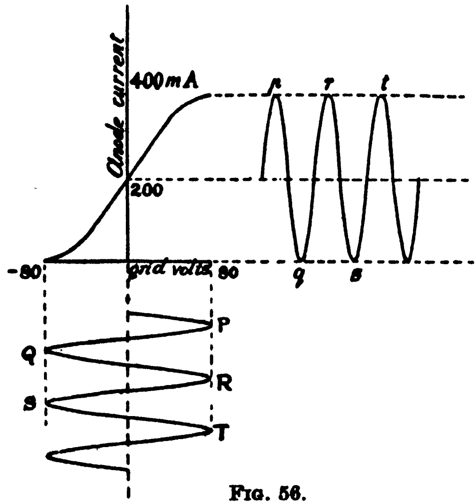


FIG. 56.

the valve. The extremes of the grid potential will then be those corresponding to the saturation and zero values of the anode current. Thus, in Fig. 56, where the *anode-current grid-volts* characteristic is represented, oscillations of current will be built up on either side of the value of the anode current (200 milliamperes) corresponding to the zero grid potential, as represented by the curve *pqrst*, while the oscillations of grid potential will be between +80 and −80 and will be represented by *PQRS*.

There is a secondary effect which must be taken into account. When the grid potential increases, we have seen that the anode current increases and the anode potential falls. But if we refer to Fig. 38, which represents a group of valve characteristics for various anode voltages, we see that the effect of a decrease of anode potential itself causes a decrease in the anode current. This means that the valve will be actually functioning on a characteristic curve giving a smaller anode current for the same grid potential. In other words, the effect of the change of anode

potential on the anode current is to lessen the increase of anode current due to the increase in the positive potential of the grid. Similarly, the decrease in anode current due to the grid potential becoming negative will increase the anode voltage, and this will itself tend to lessen the decrease in the anode current. Thus the oscillations of anode potential will tend to cancel those same oscillations of anode current which cause them.

The general effect of the drop in anode current due to increase of anode voltage will be to impose a limit on the maximum current which can be developed in the aerial or in any other equivalent load. If we turn again to Fig. 55, we see that the portion L_1 of the aerial coil is both in the aerial and in the anode circuit. The back E.M.F. of self-induction due to L_1 balances the oscillatory voltage between the anode and the filament. This back E.M.F. is given by $\omega L_1 I_a$, where I_a is the aerial current. If this voltage becomes greater than the voltage applied by the high-tension battery, the anode will become negative to the filament and the anode current will cease to flow. If this happens, the aerial oscillations cannot be maintained. Thus if V is the high-tension voltage of the point of connexion on the aerial coil (or the anode tapping-point, as it is called) this voltage must be such that

$$V > \omega L_1 I_a.$$

This gives a limit to the aerial current obtained with any given anode voltage. The correct position of the anode tapping-point is easily found in practice by varying it until the maximum aerial current is obtained.

The frequency radiated from the aerial depends on the careful adjustment of the circuits. It will change slightly with the variation of a number of factors. In general, it may be stated that the wavelength is decreased when: (1) the steady grid voltage is decreased; (2) the anode voltage is decreased; (3) the coupling of the grid circuit and anode circuit is decreased; (4) the aerial damping is increased; and (5) the filament current is increased or decreased from a value which corresponds to the minimum frequency.

When longer wavelengths are to be radiated, the grid circuit is often tuned by connecting a condenser across the grid coil L_2, so that the principle of resonance can be used in this circuit and the potential in the aerial circuit be more easily built up.

If the grid circuit is not so tuned, a large untuned coil must be used to get the necessary closeness of coupling. With close grid coupling the damping applied by the grid circuit will absorb a certain amount of power from the aerial circuit.

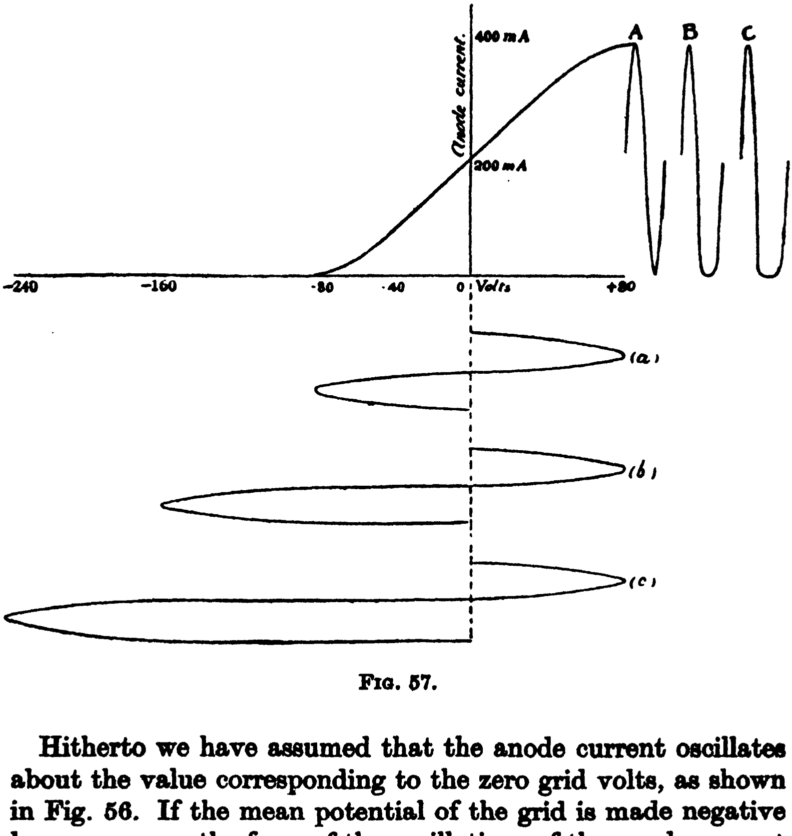


Fig. 57.

Hitherto we have assumed that the anode current oscillates about the value corresponding to the zero grid volts, as shown in Fig. 56. If the mean potential of the grid is made negative by any means, the form of the oscillations of the anode current will change. For example, in Fig. 57 curve (*a*) corresponds to an oscillation of grid potential between +80 and −80 volts, about zero as the mean value, while curve *A* gives the corresponding oscillation of anode current. In the case of curve (*b*), the mean steady grid potential is −40 volts, and the oscillations of potential take place between +80 and −160 volts, while curve *B* gives the corresponding oscillation of anode current. In curve (*c*) the grid oscillations take place between +80 and −240 volts about a steady potential equal to −80

volts, and curve C gives the corresponding oscillation of anode current.

When the oscillations given by A and C are compared, we see in curve C that current is flowing through the valve for a much shorter time than in curve A, so that its heating effect on the anode is decreased. Also the change in anode current is much sharper, so that the change in anode voltage is greater and a sharper shock is applied to the aerial circuit. The state of things given by curve B is intermediate between that given by A and C.

With the conditions given by curve A the efficiency of the anode circuit of the valve is about 50 per cent., while in the case of C it may rise to 70 per cent. The increased efficiency is partly due to the fact that less power is absorbed in the grid circuit, for we have already seen that when the grid potential of the valve is negative no current can flow between the grid and the filament. The increased efficiency, however, is realized at the expense of the purity of the radiated wave, for the wave form will become more complicated and harmonics will be present which may cause interference with other communications. We shall discuss this effect more fully later.

The question now arises as to how the grid can be maintained at the mean steady potential at which it is desired to function. One method is to insert a potentiometer between the grid coil and earth. This method is open to obvious disadvantages. For example, if through any defect the potentiometer ceased to function, the oscillations would probably cease and the full saturation current would flow into the anode, which would probably be melted. The more usual method is to make use of the grid current, and to insulate the grid from the filament and earth by a condenser shunted by a resistance. If we refer to Fig. 37 we see that when the grid is at a positive potential, a grid current flows through the valve, while when it is at a negative potential no such current flows. If the grid is completely insulated by the condenser, a negative charge will be collected, and the grid will be maintained at an increasing negative potential. If, however, a leakage path is provided by a high resistance across the condenser, this charge will leak away through the resistance, and by adjusting the rate of leak (by altering the value of the resistance) the grid can be held at a steady potential. The value of this potential will be given by the voltage drop in

the resistance; and thus, by a suitable design of grid leak, the grid can be held at any negative potential desired.

It can be shown theoretically that the maximum output of the valve is $VI_s/4$, where V is the high-tension voltage applied to the anode and I_s is the saturation value of the filament emission current. It is clear that the output can be theoretically increased without limit by raising the voltage of the high tension V. In practice a limit is imposed by the heating of the anode, and by other factors.

We have seen that to get the necessary power from a transmitting valve high anode voltages are employed. The electrons from the filament will therefore be striking the anode in large numbers and with high velocities. Under this bombardment the anode will quickly become red-hot. Hence it must be of a size to enable the heat to be radiated away, or some means must be provided for cooling it. Also, at high temperatures metals are liable to give off traces of gas, which may cause the vacuum of the valve to deteriorate. For this reason the metal parts of the valves are usually made from metals with high melting-points, such as molybdenum or tungsten, or even from specially prepared carbon, which do not occlude gas. In the same way, the bulb of the valve will also become very hot. Ordinary glass expands and contracts when heated and cooled, and is therefore liable to crack. Hence high-power valves are sometimes made from silica, which has a very small coefficient of thermal expansion. In order to obtain high transmitting power, however, a usual method is to use groups of transmitting valves in parallel.

If a number of precisely similar valves, say N, are connected in parallel, and if careful adjustments are made for each valve separately, the output of the whole group is N times that of a single valve. But, through unavoidable small dissimilarities in the valves and the difficulty of making the large number of careful adjustments necessary, it is not possible in practice to obtain this full theoretical efficiency.

For example, at the Post Office high-power station at Rugby a bank of some 52 valves is used, producing an input power into the aerial of about 500 Kw. The valves used at Rugby consist partly of glass and partly of copper. The copper portion forms a cylindrical anode which is water-cooled. Each valve is capable of dealing with 10 Kw. of energy. The construction of these

valves, which are made in Great Britain by the Western Electric Company (known as Standard Telegraphs and Cables since 1926), presents some interesting features. In sealing the copper to glass the parts are brought into contact while hot, the temperature being high enough to cause the glass to wet the metal. The metal in contact with the glass is made exceedingly thin, so thin in fact that the stresses in it when the glass cools are not sufficient to break it away from the metal. Such seals are stated to be very satisfactory, and are said to have been used from the temperature of liquid air to that of molten glass without deterioration. The principle used in the seal is also used in carrying the leads to the electrodes through the glass. To get rid of occluded gases, the metal parts of the valves are subjected to a previous heat treatment in a vacuum furnace. The valves work at an anode voltage of 10,000 volts. The average filament emission current is 1·35 ampere, but the maximum electron current which can be supplied by the filament is as much as 6 amperes. The figures given above were those in use when the transmitter was designed in 1927, and form a good example of the use of valves in parallel to generate high power. More recent designs do not employ self-oscillating valves, as will be explained. Also there is a tendency to employ a somewhat smaller number of valves each capable of handling still larger power.

We have now discussed the use of the valve as an oscillator producing radio-frequency power and directly connected to an aerial, thus forming the simplest possible type of valve transmitter. This arrangement formed a great advance over the earlier Poulsen arc or alternator transmitters, being more stable, more efficient, less costly in both installation and maintenance, and capable of satisfactory output at considerably lower wavelengths. In one major respect, however, it falls short of modern requirements, namely that of frequency stability. The radiated frequency is determined, as has been pointed out, by the constants of the resonant circuits, but is also appreciably affected by the valves themselves, their temperature and that of other parts of the equipment, and by slight changes in any of the operating voltages. Practically, therefore, a limit exists to the accuracy with which the frequency of a self-excited or self-oscillating valve transmitter can be maintained at its nominal value. This was unimportant in the days when only a few stations were in use and frequencies were low, but with the enormously greater

number of services required to-day, the constancy of something like 1 per cent. to be expected from such transmitters is not good enough, and new methods have been evolved. Another defect of the self-excited transmitter is that whilst it is reasonably satisfactory for slow-speed Morse telegraphy by hand, the radiated frequency cannot be maintained perfectly stable when telephony or high-speed automatic telegraphy is used, and this 'wobble' in frequency would cause interference to other transmitters working on very closely adjacent frequencies.

For these reasons the use of a simple oscillating valve directly connected to an aerial has been abandoned for commercial point-to-point services and the majority of other uses. It still finds a place, however, in portable and field equipment, or wherever weight and bulk are more important than a high-quality signal. Aircraft and marine emergency transmitters are typical examples of such cases. At ultra-short wavelengths also, where it is difficult to produce efficient transmitters of any other type, plain oscillators find a useful field still open. Below a wavelength of 10 metres transmission is usually intended for local communication and the waves very seldom travel more than a few hundred miles at most. Moreover, an extremely wide-frequency spectrum is available which allows transmitter frequencies to be amply spaced. Interference then ceases to be a very serious problem, and the self-excited oscillator is widely employed at the present time, although even in this field it may be expected to die out within a few years. We have dealt at some length with the oscillating valve, however, because whilst it is no longer connected directly to the aerial it still forms an essential of even the most advanced transmitters. The essential difference is that, instead of attempting to generate all the radio-frequency energy needed in an oscillator, we now content ourselves with quite a moderate power output, of a few watts only, taking special steps to ensure absolute stability of frequency. This is much easier when the oscillator is at low power and not connected to an aerial. It is then termed a 'master oscillator', and is followed by one or more stages of valve amplification which build up the power until it is sufficient for application to the aerial. The actual power valves which feed the aerial are not unlike those we have been discussing, but now they are not provided with reaction to an extent which will cause oscillation. Instead they receive an input to the grid circuit from the original

oscillator, which they merely amplify, and they may be said to be driven into oscillation by the preceding valves. For this reason the whole transmitter is often termed a 'driven transmitter'.

Before proceeding to an examination of driven transmitting circuits, however, it will be desirable to consider the steps taken to obtain satisfactory oscillation from valves at short wavelengths, between 1 and 100 metres, corresponding to frequencies of from 300 to 3 megacycles or less, and which form the most important long-distance radio channels. The circuit of Fig. 52 is an excellent one when the frequency required is low, and there is nothing in the theory of oscillation which we discussed concerning it that differs whatever the frequency may be. It is from practical causes only that departure from this simple circuit is necessary. For a valve to oscillate and deliver appreciable power into the anode resonant circuit two conditions are essential. The first of these is that sufficient coupling shall exist between anode and grid circuits; the second that the anode circuit shall be capable of resonance at the desired frequency. To these obvious requirements might be added the fact that, unless the impedance of the anode circuit is comparable with the alternating-current resistance of the valve, very little of the power generated will be usefully transferred to the external load. At low frequencies the circuit shown clearly fulfils all requirements, but what will happen if an attempt is made to work it at considerably higher frequencies?

To raise the resonant frequency, or reduce the wavelength, of a circuit it is necessary to reduce either its inductance or capacity, or both. That means that in the typical circuit of Fig. 54 (*a*) the inductance L_1 and the condenser C_1 must be reduced. Up to a point depending entirely upon the physical dimensions of the parts this will be quite simple. There is a limit, however, to the reduction of capacity, for when C_1 is reduced to the vanishing-point, capacity still remains in the form of three components, the inter-electrode capacity of the valve, the self-capacity of the inductance coil L_1, and the capacity of the aerial or external load. These can clearly be minimized, but there is a practical limit to the reduction possible in internal valve capacity, since the electrodes must have some size if they are to dissipate appreciable heat. Also the circuit connexions will have some unavoidable capacity. It is convenient to retain a small

capacity C_1 as a rule, both to facilitate adjustment and to prevent all the radio-frequency current from passing through the valve capacities. To allow this tends to set up serious heating at the 'pinch' or point where the valve leads pass through the glass envelope, and in the valve-holder. Thus we see that a very real limit exists beyond which the effective capacity acting in parallel with L_1 cannot be reduced in any given case.

To reduce the frequency a step further we are compelled to decrease the inductance L_1. Now, if reaction is by inductive coupling, a decrease in the inductances of L_1 and L_2 must limit the maximum mutual induction between them, and hence the maximum rate at which energy can be transferred from anode to grid circuit. As L_1 becomes very small, therefore, a time must arrive when it is physically impossible to obtain sufficient reaction coupling to set the valve into oscillation, and long before this condition is reached the coupling will be too weak to permit of the most efficient generation of oscillatory energy. There are two avenues open to us whereby a further reduction is possible. One of these is to combine L_1 and L_2 so that they form two portions of the same coil, and thus have the maximum possible mutual inductance. The second is to resort to capacitive reaction coupling. In both cases the circuit will be rearranged to minimize all stray sources of capacity which can appear across the inductances, and to eliminate sources of unnecessary damping which would tend to weaken the reaction effect. One important step that can be taken is the use of two valves so arranged that their inter-electrodes are in series across the inductance, the effective capacity being thus halved if both valves are identical. This form of connexion is termed 'push-pull', and will be frequently referred to later.

The simplest way to study oscillatory circuits suited for the shortest wavelengths will be to glance at a selection of those most widely used in practice, and in Fig. 58 several of these are shown. Many short-wave valve-oscillator circuits owe their inception to the pioneer investigations of amateur workers and bear the name of the experimenter who first directed attention to them. Thus the circuit of Fig. 58 (a) is the Hartley, and exemplifies the use of a single coil for the equivalent of both L_1 and L_2. We will call this single inductance L, tuned by a condenser C_1 as before, the frequency being determined by the whole inductance L in parallel with C_1 and the stray capacities.

Reaction coupling is still inductive and a choke feed to the anode is used just as in the original (Figs. 54 and 55); but in place of a separate anode and grid coil coupled by their mutual induction, two portions of the single inductance L are used. The valve cathode is taken to a point near the centre of the coil termed the cathode tap, and the portions of the inductance between this point and the two ends of the coil form in effect the anode

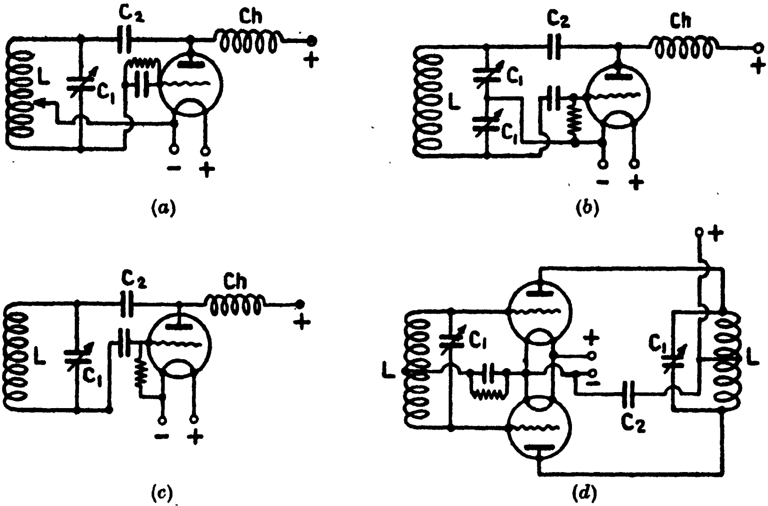


FIG. 58.

(a) Hartley. (b) Colpitts. (c) Ultraudion. (d) T.P.T.G.

and grid coils respectively, coupled by their mutual inductance as before, but more tightly than is readily done with separate coils. A blocking condenser C_2 is inserted in the anode lead to prevent a short circuit of the high-tension voltage through the coil to cathode, and when it is desired to control the reaction coupling in reception this condenser may be variable. In this way its reactance can be adjusted so that only just sufficient energy passes from the valve to the anode coil to overcome damping losses, continuous oscillation can be prevented, and the circuit becomes suitable for reception. The diagram shows the cathode tap joined to one side of the filament or to the cathode, as would be done for battery filament-heating or an indirectly heated valve. Large transmitting valves have often alternating-current-heated filaments, in which case the cathode tap would be taken to the centre point of two equal condensers joined to

each side of the filament, and also to the negative high tension or a centre tapping on the filament transformer winding. These two condensers provide a radio-frequency path of low resistance, symmetrical with respect to the filament, a condition which we have already seen is necessary for the prevention of hum.

The operation of this circuit can be regarded in a slightly different way. We have shown in Chapter VI that the grid and anode potentials of the valve are 180° out of phase at any instant, provided that the mean current supplied to the anode is maintained steady by a choke of high inductance. It is also a fact that the potentials at opposite ends of any resonant circuit oscillating at its natural frequency are 180° out of phase. Hence, if we regard the circuit LC_1 as a potential coupling between anode and grid, the anode will supply energy to this circuit, and the end attached to the anode must have a potential (with respect to cathode, for example) the same as that of the anode at any instant. But a reversal of phase by 180° occurs in the resonant circuit and so the potential at the opposite end is in opposite phase at that same instant. This, however, is the correct phase which must be applied to the grid to reinforce the original anode potential, and hence the potential transfer is such as to produce an energy increase and favour reaction.

Regarded from the potential standpoint, therefore, it is easy to understand how the Colpitts circuit shown in Fig. 58 (*b*) produces reaction and oscillation. All the remarks concerning the Hartley circuit apply equally to it, the only difference being the means chosen to provide a cathode tapping-point. It is only necessary to connect the cathode to a point on the resonant circuit intermediate in phase between anode and grid, and this can be done equally well by tapping the condenser C_1. This is done in practice by replacing C_1 by two similar condensers in series, each marked C_1 in diagram (*b*), and to connect the cathode to their common centre point. This in effect divides the resonant circuit into two parts once again, and by varying the relative capacity of the two condensers we have a control of the tapping-point equivalent to that of the preceding circuit. It makes no essential difference if the resonant circuit be divided into two portions by tapping the coil or condenser, provided that a means is available of separating the impedance it represents into two parts to form anode and grid circuits. The choice is largely a matter of convenience.

Regarded from the potential viewpoint, however, this cathode-tapping makes little difference to the mechanism of reaction, and it has been found that in the case of very high frequencies it can be dispensed with. The circuit then becomes that of Fig. 58 (*c*), which is known as the Ultraudion and has largely replaced the Colpitts because of still greater convenience. Here the resonant circuit is found between anode and grid, no cathode tap of any kind being provided. The circuit oscillates readily and is particularly serviceable at very high frequencies.

The Ultraudion does not actually operate without a cathode tapping-point, for at high frequencies the grid-cathode and grid-anode capacities within the valve provide a pair of condensers of sufficient capacity to take the place of the two condensers shown as C_1 in the Colpitts circuit. It can be regarded as virtually a Colpitts oscillator in which the internal valve capacities are relied upon to complete the circuit, and for this reason does not work well at lower frequencies, where these capacities do not offer a low reactance to the high-frequency currents.

It will be noticed that in all circuits a grid condenser and leak are shown. These play no direct part in the oscillatory circuit and merely serve to provide grid bias by the potential drop due to grid current flowing through the leak, as has been previously explained. Grid current will always exist in a self-oscillating valve circuit, and can be visualized as the high-frequency current flowing in the grid circuit rectified by the diode action between grid and cathode. The bias will therefore increase with the amplitude of oscillation, and tends to be self-adjusting to an equilibrium position. This is convenient and leads to frequent use of the grid-leak arrangement; but it must be remembered that it is not the only possible one. The leak must always connect grid to cathode, in which position it is often shown. Where a direct path to cathode exists through the coil, the leak may be connected across the grid condenser, as shown in Fig. 58 (*a*). A further point to notice is that the anode-blocking condenser which is essential in the Hartley and Ultraudion circuits could be omitted from the Colpitts or T.P.T.G. which follows, but may be retained as a method of controlling reaction when needed.

None of the first three circuits mentioned employs true capacity reaction coupling, and in Fig. 58 (*d*) is shown one that does this. This is the Tuned-Plate Tuned-Grid circuit, so called for want

of a better name, and often abbreviated to T.P.T.G. It is shown in the push-pull form as an illustration, and it should be clearly remembered that each of the circuits shown can be used either as a single-valve arrangement or duplicated 'back to back' to permit of the use of two similar valves in push-pull. The present circuit employs a simple resonant circuit between grid and cathode of each valve, and also between anode and cathode, but no magnetic coupling is allowed between the two circuits. Coupling occurs simply through the anode-grid capacity of the valves, which forms a small condenser between the anode and grid resonant circuits, and in the case of most valves transfers sufficient energy to produce oscillation. Clearly the circuit is dependent upon valve characteristics, some types being more suitable than others, whilst a few have insufficient inter-electrode capacities to perform in it at all; also the coupling effect of a small capacity will naturally increase at high frequencies, making the arrangement best suited to these. As an oscillator nothing more need be said concerning this circuit arrangement, which is both efficient and widely used; but the advantages of push-pull working merit attention. It will be seen that the electrode capacities of the two valves are in series across each resonant circuit, and their effect in limiting the maximum frequency to which the circuits will tune is minimized. Secondly, the circuit is symmetrical about the cathodes, and, since these are usually earthed, the high-potential points are symmetrical with regard to earth. This is helpful in many ways, as it assists stability and minimizes all kinds of leakage fields and potentials which can upset the operation of near-by equipment. It is now possible to feed the anode circuit with direct current from a centre tapping on the anode inductance, a blocking condenser C_3 preventing a short circuit to cathode. This point will be at cathode potential, however, for radio-frequency currents, and this prevents radio-frequency energy from leaking into the high-tension supply circuits, often a trouble in the single-valve arrangements. Since the point is already one of zero potential-variation it is possible to omit the choke entirely, for there is no need to hold steady an already uniform current. This is a decided simplification in practice, for energy losses are unavoidable in this choke at high frequencies. These advantages of push-pull operation will become more obvious when we come to study amplification.

These four circuits furnish good examples of the modifications that enable valves to oscillate at shorter wavelengths than were originally possible. There are of course a large number of slighter modifications which do not affect the underlying principles, and a wide variety of circuits differing only in minor details. The highest frequencies obtainable do not depend only upon the circuit, however, but also upon the design of suitable valves having low inter-electrode capacities and high mutual conductance, and upon the physical construction of the components employed. In the construction of valves for other uses it has been general practice to bring out the anode, grid, and all other leads to a single base which can conveniently fit into a single socket. Modern valves intended specially for very high-frequency working no longer do this, only the filament, cathode, and screening-grid leads being brought out to the base. Anode and grid leads are brought through the bulb separately at widely removed points so that the internal leads are short and well clear of all others. They may be led out through two horns at the opposite end of the bulb, as indicated by the sketch in Fig. 60 on page 145. This cuts down the stray capacities to an absolute minimum, and makes it possible for such valves to oscillate in the type of circuits shown at frequencies up to about 300 megacycles. Older types having all leads brought out at the base may not oscillate below 50 to 100 megacycles, and even then with lower efficiency.

Efficiency at high frequencies largely depends upon the design of all the component parts which make up the circuit, and particularly upon the inductance L and the condenser C_1. Actually the energy dissipated in heat tends to become greater, in the case of most components as frequency rises, and such sources of loss as radio-frequency resistance, dielectric losses, and contact resistance become very much greater. All parts such as valve-holders or blocking and by-pass condensers must be specially designed to minimize these losses. Special insulating materials have been developed having exceptionally low dielectric losses which do not increase seriously in high-frequency fields, since the older dielectrics such as ebonite or bakelite have been found to be very poor in that respect. The new materials are synthetic and have various trade names, Isolanite, Frequentite, Ceramic, XP-53, Victron G, Vitrolex, and Micoroid being typical examples, mostly of American origin. It is also necessary for all

leads both internal and external to be short, and metal parts of components which carry high-frequency current must be small and rigid, since a few inches of conductor can offer an impedance of several hundred ohms in the ultra-high-frequency spectrum. The impedance and inductance of these leads is not of so much importance from the aspect of lost power, but may cause undesired coupling if it be common to two circuits which should be independent, and can also raise the potential of components above 'earth' when this is not desired. It is usual to regard a certain datum line in all transmitting equipment as at zero potential, and to connect it to 'earth' in order to prevent any variation in potential. This is frequently the negative high-tension conductor, to which are bonded the valve cathodes, low-potential ends of resonant circuits, metal screening-compartments, and all portions of the circuit and apparatus intended to be at constant high-frequency potential. Even when the earth forms no part of the radiating aerial circuit it is necessary to have all or most of these points at earth potential, mainly to ensure stable working of the equipment and freedom from body-capacity disturbances by those operating it. At low frequencies it is quite satisfactory to bond all such points to the nearest convenient earthed metal framework or chassis, but at very high frequencies the inductance of even short leads makes this unsatisfactory. The alternative is to return all circuits associated with each particular valve stage to a common point, which is then joined to other nominally earthed points and to the earth itself by a single thick insulated conductor. This point is mentioned here so that the meaning of the expression 'the earth line' will be understood, and the reason why special steps are necessary to get true constant-potential points in very short-wave apparatus. Most radio-circuits only work correctly on the assumption that some point, usually the valve cathodes, is at constant radio-frequency potential.

The inductance and condenser forming the resonant circuit are also capable of improvement. When operating at long wavelengths the inductance can consist of copper wire, strip, or tube, or may be wound of stranded wire in which each strand is individually insulated. It may be supported by an insulating former or framework, one of the best materials being thoroughly dried wood. Above a frequency of about two megacycles stranded conductors provide no decrease in high-frequency

resistance, and their use is not justified, the favoured conductor being copper tube or heavy-gauge solid wire. This is supported when necessary on a former built from one of the low-loss materials mentioned in the preceding paragraph. High-frequency currents travel mainly in the surface layer of the conductor, and this 'skin effect' makes it necessary for the surface to be highly conductive. A copper surface is therefore kept clean and often burnished, whilst it is protected from oxidation by a coating of varnish or enamel. A still better process is to silver-plate the conductors, which is the usual procedure at frequencies above about 30 megacycles. In the very high-frequency region the inductance may consist of a few turns of silver-plated copper tube, of small diameter and rigidly self-supporting without the help of a former. An inductance of this kind remains efficient up to the highest frequency at which triode valves will oscillate in reactive circuits.

The condenser is of equal importance but less capable of modification. It will be of the familiar multi-plate construction, and at long waves may employ oil or mica as a dielectric, since the high dielectric constant of such materials assists in obtaining the necessary high capacity. At higher frequencies the best possible power factor and minimum dielectric loss in the intense field between the plates make air the only permissible material, whilst at low frequencies it is also the best dielectric to employ whenever possible. By-pass and blocking condensers employ mica dielectric even at the highest frequencies, but precautions must be taken to see that the conducting path between the plates is short enough to have negligible inductance. Recently it has been found that certain of the synthetic plastic materials, such as Ceramic, are also very suitable as dielectrics. It is possible for the inductance of condenser electrodes to raise the impedance of the whole, so that even a large capacity will not offer the small impedance that is expected. We may get the curious anomaly of a condenser of larger capacity exhibiting an increased impedance owing to the greater inductance of the electrodes and their connexions, and therefore performing poorly. For any chosen construction there will be an optimum capacity best found by measurement to give tne minimum impedance at a specified high frequency. The condensers must be as non-inductive as can possibly be attained. Contact resistance must be guarded against, since it will represent an

increased proportion of the whole impedance, now a very low figure, and the plates must be brazed together rather than bolted. The actual resistance of the plates themselves may be no longer negligible, and there is a tendency to silver-plate these when the frequency exceeds about 30 megacycles. It is often possible to dispense with the condenser C_1 entirely and to rely upon the valve and stray capacities, provided that the power involved is not so great that the whole high-frequency current flowing through the valve seals will cause damage through overheating. At high powers it is necessary to polish condenser plates and to round off all corners or sharp edges. This is done to prevent brush discharges or flash-over at high voltages, and is as important at low as at high frequencies, because voltages are likely to be greatest in the former case.

Mechanical rigidity of the inductance and condenser is one of the important factors that make for stability. Obtaining the minimum possible circuit losses also results in a circuit which tunes sharply and is less prone to change in frequency, and one of the two principal methods of stabilization depends upon this property. The only method of stabilizing the frequency of a simple self-oscillating valve transmitter working at a very high frequency is to sharpen up the tuning in this way, since the 'fly-wheel' action of the resonant circuit in holding the whole transmitter to a definite frequency increases as its damping is reduced. Now, any circuit will resonate provided that it has inductance, capacity, and low damping, and we shall see when discussing resonant aerials that even a length of wire suspended in space has a very definite resonant frequency at a wavelength twice the length of the wire. There is no absolute necessity for inductance and capacity to be separated into two individual units, such as the coil and condenser of long-wave circuits, and we have just seen that many factors make this impossible when the wavelength becomes extremely short. Condensers have some unavoidable inductance, and the inductances have self-capacity. These can be replaced by a resonator having 'distributed' inductance and capacity combined within a single component, and it is found very useful to do this at wavelengths below about 10 metres. We speak now in terms of wavelength because when capacity and inductance are distributed it is simpler to relate the physical dimensions of the radiator to the wavelength than to think in terms of frequency.

A number of physical structures which behave as resonators can be conceived, and in most cases the rigidity and consequent frequency stability is considerably better than that to be expected from conventional coils and condensers, whilst the radio-

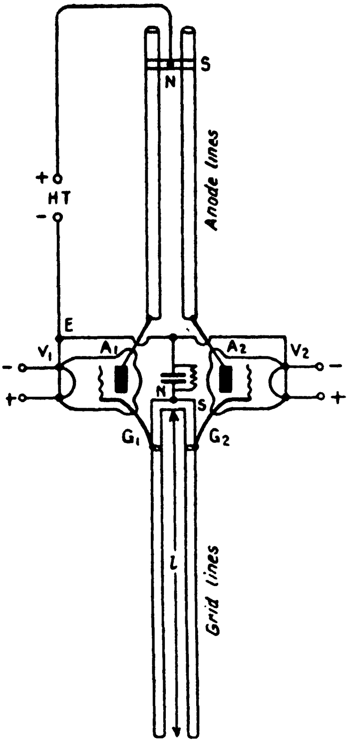


FIG. 59. Linear Oscillator.

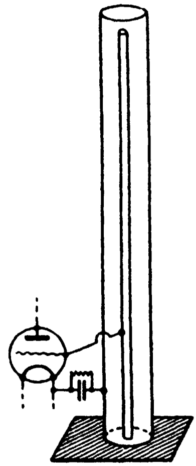


FIG. 60. Concentric Line Oscillator.

frequency resistance and losses are much less. When discussing selectivity later a factor Q will be introduced. This is a figure the magnitude of which measures the 'goodness' or low damping inherent in a tuned circuit, and it is usual to describe these resonators as 'high-Q' circuits. Their importance in ultra-short-wave transmission is sufficiently great to justify a description of two of the more important types. These are illustrated in Figs. 59 and 60.

The so-called 'long-lines oscillator' of Fig. 59 exemplifies one construction in which the resonant circuit consists of two parallel

copper tubes. These may be silver-plated to reduce surface resistance. The tubes are bridged across at one end by a heavy copper strap *S*, so that the whole resonator becomes a copper tube bent into the form of a long narrow 'U' or hairpin. The advantage of a strap is that it can be slid along the tubes to vary the effective length of the 'U', which determines the wavelength at which the device resonates; but when no adjustment is wanted the strap is often omitted and the whole built from a single tube bent into hairpin shape. The inductance in this case is that of the loop of copper tubes, and the capacity that between them, both being distributed along the tubes and not 'lumped' as in a coil-and-condenser arrangement. The circuit shown illustrates a push-pull oscillator employing linear resonators for both grid and anode circuits, reaction coupling being through the internal valve capacities. It is thus strictly comparable with the circuit of Fig. 58 (*d*). Notice that the use of push-pull makes the whole arrangement symmetrical, and that no anode feed choke is necessary, the high-tension supply being taken to the mid point of the anode resonator, which is at minimum radio-frequency potential. In the same way the grid condenser and leak, common to both valves, can be inserted between the centre of the grid resonator and the cathodes, points of substantially equal radio-frequency potential. There should therefore be no oscillatory energy lost in these components, as would be the case in a circuit such as Fig. 58 (*c*), in which the leak is virtually in parallel with the oscillatory circuits and forms a wasteful load across it. It is sometimes necessary to insert chokes in both filament leads if the oscillator is to be used at the highest possible frequencies, as then no radio-frequency current flows into the filament-supply leads and the filament or cathode takes up a natural intermediate potential between anode and grid. This is preferable to any attempt to earth the cathode directly, since an earth lead, however short, possesses so great an inductance that it fails to provide a low-impedance path to earth, merely forming a source of unwanted radiation and spurious effects. The diagram also illustrates the use of valves having anode and grid brought out at separate points, permitting short leads to the resonators and a reduction of stray capacities between them and all cathode or supply circuits.

A clear distinction must be drawn between two types of linear oscillator. The first is not illustrated but is very similar to the

Letcher wires used in wavelength measurement. Here two copper tubes would be used, one forming the grid and one the anode resonator of a valve. These are attached to the valve electrodes at one end and run parallel as in the illustration, but there is no connecting strap between them. Here each tube, including the lead which joins it to the valve electrode, is made exactly one-half wavelength long, and will resonate as a conductor in free space, except that the wavelength will be slightly lengthened in practice by the capacity due to surrounding objects. Anode- or grid-supply leads can be tapped on to the resonator at the potential nodal point exactly half-way along the tube. The operation of this type of circuit will be more clearly grasped after we have considered resonant aerials, to which it is closely related. The frequency stability is relatively very good, but it suffers from practical drawbacks in that the length of the resonators is large, making the equipment bulky, whilst radiation can occur from the resonator itself.

The half-wave line just described may be termed the true 'long-line' or resonant-line oscillator. It becomes more practical if modified in the manner of the grid circuit of Fig. 59. Here the half-wave line is bent into a 'U' about the central nodal point, so that the two quarter-wave sections lie parallel to each other. The overall length is thus reduced to somewhat less than half, whilst the fields of the two portions tend to cancel each other in the surrounding space, reducing radiation to a low figure. It will be noticed that the valve grids are tapped on to the resonator a short distance each side of the nodal point N. In this way a sufficient oscillatory voltage is tapped off, whilst little loading is transferred to the resonator.

The anode circuit, on the other hand, represents the second class of linear resonator, since the capacity added by the valve anodes joined across the whole inductance raises the resonant wavelength considerably, and the length of the tubes becomes a good deal less than a quarter-wavelength. No exact relationship remains between the two, since the valve anode-to-cathode capacities have become a determining factor, and the resonator is therefore termed a 'short-lines' linear oscillator to distinguish it from the former type. It still retains a large measure of the high Q associated with linear construction, and whilst it contributes less to frequency stability than does the grid resonator it assists in obtaining good efficiency from the

oscillator. The former has a very low damping coefficient (high Q), which is not reduced to any large extent by the grid connexion; the grid resonator thus contributes a high degree of constancy to the wavelength.

One further type of resonator will be described, that employing concentric tubular lines. This is illustrated in Fig. 60 and comprises a central tubular conductor slightly less than one-quarter wave in length, surrounded by a second tubular conductor of larger diameter and slightly greater length. The inner tube determines the wavelength and the outer forms a protecting screen against outside disturbances, whilst its inner surface completes the resonator. The two tubes are solidly joined together at one end only, as by brazing into a thick copper plate. This joint comes near the voltage node and carries maximum radio-frequency current. Consequently any resistance at this point will seriously reduce the efficiency of the resonator. The optimum ratio of tube diameter to spacing in all linear resonators is given by $x/y = 3{\cdot}6$, where x is the inner radius of the outer tube or the spacing between parallel tubes, and y is the diameter of the inner tube or of both when these are of the open parallel type. As an example of the possible values of Q obtainable which can be compared with those of ordinary circuits to be given later, an unloaded linear resonator employing two tubes of 3 in. diameter has at a frequency of 50 megacycles (wavelength of 6 metres) a Q of nearly 5,000, and this increases proportionally to the square root of the frequency when all other factors remain unchanged. For comparison, a Q of 300 is difficult to obtain from ordinary coil-condenser types of resonant circuit, and, since the frequency stability is broadly proportional to the value of Q, we see that a large improvement has resulted from the use of distributed constants. Other resonators employ disks, hemispheres, and a variety of metal structures, but their scale of utility is not great enough to warrant description in these pages.

Before proceeding to other methods of frequency-stabilization it will be interesting to refer briefly to the methods for the production of oscillations still shorter than those possible from reactive circuits of any kind. At a wavelength slightly below 1 metre, depending entirely upon the valve characteristics, oscillation of this type becomes impossible. This is due to an entirely new effect, namely, that the time taken for electrons

to pass from cathode to anode has become of the same order as the time period of the oscillations. When this condition is reached, the grid loses control over the oscillatory anode current, for electrons accelerated or decelerated by the grid potential never have time to reach the anode before a reversal of potential takes place. The whole operation of the valve along ordinary lines therefore breaks down. The only way in which it can be extended to higher frequencies lies in a reduction of valve dimensions, and this implies small electrodes and reduced power-handling capacity, leading to a deadlock when large power is required.

Fortunately it has been found possible to obtain oscillations from the triode valve by new methods capable of very much higher frequencies. These oscillations take advantage of the very limitation in electron-transit time mentioned above, and are electronic rather than regenerative oscillations. The first of these to be developed was that due to Barkhausen and Kurz in 1919. The valve used must be of symmetrical construction so that the path from cathode to anode is equal at all points, and in operation the grid is maintained at a higher potential than the anode. The frequency of oscillations depends mainly upon these potentials and upon the filament emission and geometry of the valve electrodes. It is believed that electrons pass through the highly charged grid at high velocity but, having insufficient energy to reach the anode against the back-pull of the grid, turn back whilst in the anode space to oscillate backwards and forwards through the grid mesh. The complete explanation of the action is complex, but in the papers of Gill and Morrell it is shown that the energy expended in causing oscillation of the electron cloud within the anode-grid space can be negative, thus producing a condition analogous to negative resistance. Energy will then be supplied to an external circuit between grid and anode, and, if this be adjusted to resonate at the same frequency, it will be thrown into oscillation. An important difference is that the external circuit plays little part in determining frequency, this being a function of the conditions existing within the valve. The lowest wavelength practicable by the Barkhausen-Kurz effect lies between the limits of 10 and 30 cm.

More recently the split-anode magnetron valve has shown up the possibility of two other types of electronic oscillation. The magnetron is a valve similar to the diode but operating in an

intense axial magnetic field generated by an external electromagnet of constant strength. The split-anode modification employs a cylindrical anode divided longitudinally into two portions. Electrons travelling under the influence of this field will describe curved paths, the curvature depending upon the electron velocity and field strength. By suitable choice of these factors electrons can be caused to describe closed paths within the anode space, and can oscillate continuously in these orbits whilst drawing energy from the field. In the scope of this work it is impossible to discuss such oscillations fully, and we must be content with recording their existence and practical utility. Those usually termed electron oscillations do not necessarily require a split anode, and resemble the Barkhausen-Kurz oscillations in that they are dependent upon internal valve conditions and not upon the external circuit resonance. A second type may be termed dynatron oscillations and occurs at a frequency determined mainly by a resonant circuit connected between the two portions of the split anode, which resemble push-pull valves to a certain extent. The latter type have been the subject of extensive work by McGaw and are capable of surprising efficiencies approaching 50 per cent. at 100 megacycles. This implies that up to half the energy supplied to the anode circuit as high-tension current is converted into radio-frequency power and greatly exceeds that possible from reactive oscillators at so high a frequency. In these and other forms of electronic oscillator we have means of generating wavelengths down to a few centimetres, and research now in progress is extending this range daily and improving efficiency near the lower limit. The above methods provide us with the highest frequencies known, but at more ordinary frequencies there is an entirely different method depending upon the selection of successively higher harmonics of a lower-frequency oscillator. When the highest stability is required the greater complexity of this method is thoroughly justified. It is the basis of most modern short-wave transmitters, and we are now ready to consider it in more detail.

The advantage of generating a radiated frequency by harmonic selection and amplification from an oscillator of lower frequency and power lies in the ability to stabilize the latter very completely. This is due to two factors: one that the oscillator can work under conditions of constant load, and its frequency need not be disturbed by the necessity of modulation or variable

aerial-load conditions; and the second that an oscillator of low power and relatively low frequency can be stabilized by methods not applicable to high-power units, and can be maintained at uniform temperature if necessary or supplied with voltages of high constancy.

The methods adopted for maintaining constant the frequency of this master oscillator fall into the following three groups:

(a) The use of 'high-Q' resonant circuits combined with controlled temperature and supply potentials.
(b) The control of frequency by a piezo-electric resonator or quartz crystal, commonly termed 'crystal control'.
(c) Control of frequency by a mechanical resonator such as the tuning-fork.

The first of these methods has been partly dealt with in the preceding description of high-Q oscillators for ultra-short-wave transmission. Exactly similar circuits to those described can be used for master-oscillator purposes, and certain of them, such as the concentric resonant line, are widely adopted when the oscillator is to be of exceptionally high frequency. In general, however, the master oscillator will work at a frequency not higher than 10 megacycles, and ordinary coil-and-condenser circuits are satisfactory. Stability is obtained by attention to circuit details and to extremely rigid mechanical construction of the components used. Circuits are mostly of the Hartley type shown in Fig. 58 (a), modified in minor ways to conform to the designer's preferences. The whole oscillator is frequently enclosed in a constant-temperature oven, thermostatically controlled to an accuracy which may be as high as 0·2° Centigrade; or when this is not justified attempts are made to design the resonant-circuit coil and condenser in one of several ingenious ways by which thermal expansion is self-compensated. Materials of equal and opposite thermal-expansion coefficients are chosen, mechanically designed so that expansion produces negligible changes in actual electrical values. Voltage supplies are derived from very constant sources such as secondary cells, and may be held to a constant value by automatic relay devices. A leading example of this type of oscillator is the Franklin master oscillator developed by the Marconi Company for use in much of their commercial equipment, and widely used in ocean-going-liner installations. An important advantage of

this system over others is that the oscillator can incorporate adjustable tuning, and can be set to any one of a number of selected frequencies by simple switching. It is thus flexible and ideal should a transmitter need to operate on a number of quite widely differing wavelengths.

The second type of controlled oscillator is that which employs a piezo-electric crystal, nearly always of quartz; and so very important has this type become that we must digress briefly to describe the principle of piezo-electric resonators. The property of piezo-electricity was discovered by J. and P. Curie in 1880, but its practical application dates largely from the work of Cady commenced in 1922. It is found that if a plate of quartz be cut from a natural crystal in certain definite planes and a mechanical stress applied to its surfaces, an electrical potential appears across these (or in some cases other) faces, and is proportional within wide limits to the applied stress. The converse effect is equally true, that an electric potential suitably applied to the quartz plate will give rise to corresponding stress, which in turn must produce a slight deformation of the crystal against the opposing force of its own elasticity. Now quartz is a highly elastic solid, and a regularly shaped plate or bar will be capable of mechanical vibration at some resonant frequency depending upon its dimensions, in exactly the same way as a steel reed or violin string. A quartz plate freely held in suitable supports which do not restrain it will vibrate very freely and with very low decrement if excited at one of its natural periodicities, forming a first-class mechanical resonator. But in doing this it is subjected to periodic stresses and must necessarily give rise to electrical potential variations at the same frequency as its own natural vibration; moreover, since the electrical and mechanical stresses are strictly and reversibly interconnected, the application of electrical potentials alternating at that frequency will set the quartz into resonant vibration just as would mechanical excitation. The increased amplitude of vibration reinforces the original potential oscillations, tending to build up to a maximum at which the applied electrical power is dissipated as heat and mechanical work at the same rate as it is supplied. The quartz plate therefore acts very similarly to an electrical resonator, and in fact behaves towards associated circuits as would a tuned circuit of very low damping. Fortunately the resonance of the quartz is so sharp that it behaves as an electrical

circuit of extremely high Q, unattainable by the use of coils and condensers, and the frequency is rigidly determined by physical dimensions capable of extreme accuracy. It therefore provides an excellent method for controlling the frequency of a valve oscillator, in which the crystal resonator merely replaces the conventional tuned-grid circuit.

The natural quartz crystal belongs to a very important group of substances termed bi-axial crystals, on account of their property of exhibiting different refractive indices to light-rays polarized in two mutually perpendicular planes. The transparent crystal possesses an axis of symmetry termed the 'optic axis' because rays of light parallel to this axis are differentially refracted in the above manner, and it also possesses a second principal axis, perpendicular to the optic axis, termed the 'electric axis'.

A number of piezo-electric substances are known, that showing the strongest piezo effect being Rochelle salt, or sodium potassium tartrate. This material is soft and not capable of accurate grinding to the exact dimensions necessary for frequency control, so, whilst it is finding wide application in the construction of microphones and loud speakers, quartz is universally chosen for the former purpose. A second mineral, tourmaline, is sometimes used in place of quartz, since it yields a larger crystal for a given resonant frequency, and consequently can be used at higher frequencies without risk of breakage or production difficulties. It has not yet found wide application, however, as it is more costly and more difficult to cut than quartz, whilst the latter is quite satisfactory in the great majority of cases. Good specimens of tourmaline are comparatively rare.

Natural quartz occurs in hexagonal crystals, usually comprising a hexagonal prism terminated at each end by a pyramid, and the axis joining the vertices of these two pyramidal extremities is the principal axis of symmetry of the crystal. It is also the axis along which quartz exhibits its remarkable optical properties, as a ray of light passing along it emerges with the plane of polarization rotated through an angle proportional to the thickness of quartz through which the ray has passed. The optic axis is most important electrically, since it is the only direction in the crystal along which piezo-electric effects do *not* occur, whilst they are at a maximum in directions perpendicular

to it. When preparing a quartz plate for oscillator purposes it is usual to cut a slice from the crystal perpendicular to the optic axis, thus obtaining a hexagonal plate. This hexagon is not regular but has three sides longer than the other three, and therefore it has not six but only three axes of symmetry, each of which is perpendicular to the optic axis. Each of these is termed an electric axis, since an electric field acting parallel to it produces the maximum piezo-electric effect.

Before cutting an actual resonator from the quartz plate it is essential to select a portion of uniform electrical properties. Quartz exists in two forms, the right-handed and left-handed crystal respectively, having the property of rotating the plane of polarization to the right or left, and having piezo-electric coefficients of opposite sign. Unfortunately, these two varieties often occur mixed in a single specimen, but a resonator containing portions of both would be either of poor efficiency or totally inactive. Hence the sample of quartz is examined under polarized light, in which a uniform specimen looks uniform in colour whilst impure portions show striking colour-variations. A uniform section free from any kind of flaws is selected, and may be further tested for piezo-electric activity, after which a resonator slightly greater than the required size is cut from it. Quartz is a very hard material and is cut laboriously with the help of carborundum powder moistened with water and glycerine and applied to the edge of a steel or brass tool. It is important for this reason not to cut without ample preliminary testing. The resonator, having been cut out, is ground with great care to the exact size desired, the frequency being experimentally tested from time to time. The surfaces must be exactly parallel, as if this is not so there will be a tendency for parts of the quartz to resonate at differing frequencies rather than as a whole. This lowers the efficiency of the resonator, and introduces secondary frequencies which spoil it for many uses. The surfaces must be plane and smooth but it is found better not to give them a high polish. The success of the finished resonator depends largely upon these two factors of accurate grinding and perfectly uniform quartz.

A very usual form of resonator is the rectangular bar or plate. This is cut from the hexagonal section so that one of its axes is parallel to one of the three electric axes. We then have a resonator in which one dimension is parallel to the optic axis of

the original quartz crystal, a second perpendicular dimension is parallel to the electric axis, and the third dimension is mutually perpendicular to these two. This last may be called for reference the 'third axis', or by some workers the 'mechanical axis', because stress parallel to it results in electrical potentials along the electric axis. For a rectangular resonator these three dimensions can be identified with the thickness, length, and width of the bar, which when cut in this manner is termed an 'X-cut' or 'Curie-cut' crystal. In use it is mounted between two metal electrodes so that these are perpendicular to the electric axis and are either in light contact with the crystal faces or separated therefrom by a very small air gap. As the several axes may be found somewhat confusing a sketch of the arrangement is given in Fig. 61.

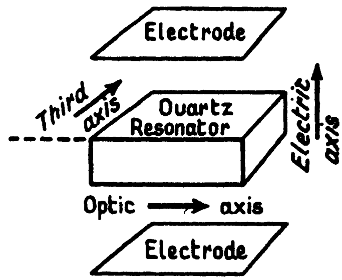


FIG. 61.

The resonant frequencies depend upon all dimensions of the crystal cut, and there will be several possible frequencies which correspond to different modes of vibration. Two of the most widely used and simplest to conceive are the longitudinal vibration of a bar resonator and the 'thickness' vibration of a thin plate. The former is used for low-frequency resonators in which the crystal is cut to the form of a long rectangular bar, and the frequency depends mainly upon the length parallel to the third axis. Radio-frequency oscillators are of the second type, in which a thin circular or rectangular plate vibrates at a frequency determined by its thickness, the dimension parallel to the electric axis. An almost infinite variety of possible cuts exists, and a very large number of vibration modes can be excited. These differ in their properties and find widely different applications, whilst many of them are still the subject of research.

We shall have occasion to refer to some of the other applications of quartz when dealing with modern selective circuits, but for the present purpose the principal property of interest is the temperature coefficient, or the rate of change of the oscillation frequency per degree change of temperature. A simple X-cut plate was considered satisfactory as a transmitter control for many years, but its frequency changes by about -40 cycles in a million per degree Centigrade, the negative sign indicating

that the change is a decrease in frequency for increased temperature. This means that for very accurate control the frequency will drift as the quartz warms up under excitation, or will vary from day to day with the weather unless the resonator be enclosed in a constant-temperature oven. Its importance will be better appreciated by considering the example of a carrier wave

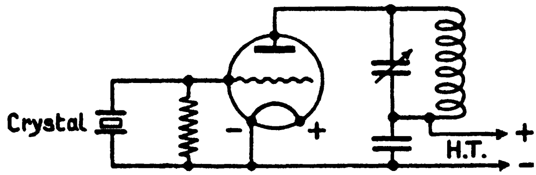


FIG. 62 (a).

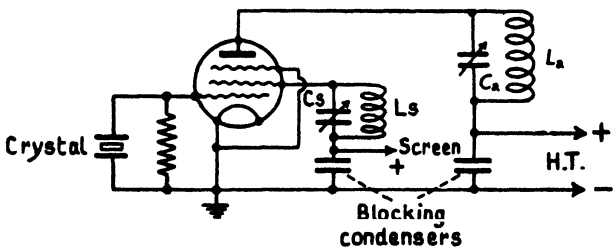


FIG. 62 (b).

of frequency 30 megacycles, which would be changed by 12 kilocycles if the crystal were heated 10 degrees, a sufficient change to lose the signals completely in a selective receiver. Recently it has been found that the temperature coefficient can be greatly reduced by modified forms of cut. One method makes use of the fact that a crystal cut with one dimension equally displaced between two of the electric axes instead of parallel to one of them will possess a positive temperature coefficient, being sometimes termed a 'Y-cut' plate. It would be expected that in between these two cuts, having negative and positive temperature coefficients respectively, might be found one of zero coefficient, in which the two types might be regarded as equally and oppositely present. This is in fact the case, and resonators are now prepared in which the active dimension makes an angle of some 30 degrees with the electric axis, and which have an extremely low coefficient. These are termed 'zero-temperature-coefficient' or sometimes 'skew-cut' resonators, and at the

principal resonant frequency temperature-variation is negligible, an oven being unnecessary for the majority of applications.

Having now an idea of the preparation and properties of crystal resonators we can examine the manner in which they are used to control the frequency of valve oscillators. The simplest circuit is shown in Fig. 62 (a), and will be seen to resemble the tuned-plate tuned-grid oscillator but with the crystal in place of a tuned-grid circuit. In operation it is very similar also. Fig. 63 shows the equivalent electrical circuit or

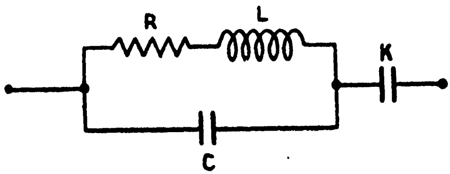


FIG. 63.

network to which the quartz crystal corresponds, and it will be seen that both 'series' and 'parallel' resonance are possible. Both mathematically and physically the crystal behaves as a combination of inductance, resistance, and capacity shown, provided that the frequency is close to one of the principal resonant frequencies of the crystal. The equivalent values of L, C, R, and K (the latter including the holder capacity) are not the same, however, for other modes of oscillation, and have no validity at non-resonant frequencies. They are values which cannot generally be realized by actual inductances and condensers, L being very large, C a few micromicrofarads only, and R lower than is possible for any practical inductance except at very high frequencies. At lower frequencies, therefore, the damping is less than for a normal tuned circuit, and resonance correspondingly sharper. Inserted into the grid circuit of a valve the crystal resonator behaves as a tuned circuit at certain natural frequencies, and, if the anode circuit be tuned in the usual way to nearly the same frequency, reaction will occur through the valve capacity, and oscillation set in. Since the quartz behaves as a parallel tuned circuit over only a very narrow band of frequencies, oscillation is confined to these and is little affected by the valve or its operating potentials. The curve of Fig. 64 shows how the frequency of oscillation varies as the anode-circuit tuning condenser is rotated, and it will be

noticed that near the point P there is very little change indeed. Working at this point the frequency is virtually determined by the crystal only and affected by the two factors of crystal temperature and the gap (if any) between the crystal and electrodes. The latter effect is small and is sometimes used as a fine adjustment.

Under oscillating conditions energy is dissipated in the resistance of the crystal R, and considerable strains are set up within the quartz. A maximum amplitude is determined by these factors, and, if this be exceeded, the crystal will crack. The energy existing in the grid circuit must be kept down for this reason, and it is not possible to control large power valves directly from a crystal. The radio-frequency power which can be drawn from the oscillator anode circuit will clearly depend upon many factors, such as the valve rating and characteristics, the voltages used, the crystal dimensions, and the actual frequency. A high-μ valve is desirable. Assuming a valve of ample size and amply supplied, then large crystals will deliver the greatest power. Crossley has shown that in practice maximum power is obtainable at a frequency of about 3,500 Kc. and falls off slowly on each side of this figure. A minimum figure is perhaps 1 watt, but can be readily increased to some 30 watts when using ordinary receiving valves. High output is assisted by the use of a choke and battery to supply bias in place of the grid leak shown in Fig. 62. This maintains the grid more continuously negative and eliminates most of the losses caused by large grid current during the positive half-cycles. By the use of large valves and special precautions it is possible to obtain as much as 100 watts of power, but for master-oscillator purposes this is far more than necessary.

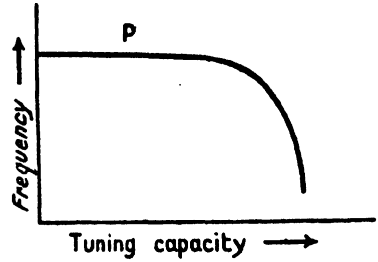


FIG. 64.

Up to the present, triode valves only have been mentioned in this chapter as valve oscillators. As a general rule no advantage arises from the use of multi-grid valves as oscillators, but the case of crystal control may be an exception. Fig. 62 (b) illustrates a type of oscillator having considerable advantages, in which a pentode valve is used with the crystal connected from grid to

cathode as before. It is possible to 'earth' the screen-grid as usual and to tune the anode circuit to the crystal frequency. When this is done the pentode behaves like a triode with the advantage that its higher amplification necessitates less oscillatory amplitude at the grid to produce a large anode power. The crystal will therefore oscillate with less amplitude, and with less risk of damage, whilst more power can be safely drawn from the pentode.

If a circuit tuned to the crystal frequency is inserted into the screen-grid lead, which then behaves as the anode of a low-powered triode, it will produce oscillation in the usual way. This oscillation must modulate the electron stream passing through the mesh of the earthed suppressor grid to the true anode, causing the anode current to oscillate also. We can now draw off energy from the anode circuit, and, since this is screened from the oscillator portion of the valve, any slight effect that varying anode loads might have upon the frequency is reduced. Since the anode circuit is only coupled to the crystal circuits by the electron stream, this form of circuit is termed an 'electron-coupled oscillator'. It works equally well in reducing the effects of anode load upon frequency, whether the original oscillation be crystal-controlled or not, and consequently is very widely used in modern receiver and transmitter design.

Owing to the effects of grid current during positive half-cycles of the oscillatory grid potential, the latter is not strictly sinusoidal. The anode current is an amplification of this and also not sinusoidal, but distorted. It can be shown that such a distorted wave form contains strong harmonics of the original frequency, and hence, if the tuned circuit $L_a C_a$ be tuned to one of these harmonics, it will be thrown into resonant oscillation at that frequency. In this way we can select any harmonic of the crystal frequency and draw off energy from the anode circuit at this new frequency. The power available will be quite a useful proportion of the total in the case of the second or third harmonic, becoming quickly smaller as the fourth and higher harmonics are attempted, but it is quite possible to obtain from 20 to 30 per cent. of the fundamental output at twice or three times the crystal frequency. This process of harmonic selection is called 'frequency-doubling' or '-multiplying', and is of the greatest importance in transmitter design. The circuit of Fig. 62 (*b*) may thus be termed a pentode crystal oscillator, if the anode

circuit be tuned to the crystal frequency and the screen-grid earthed; or as shown and with the anode circuit tuned to twice the crystal frequency it becomes a crystal oscillator doubler. It is possible to obtain very good harmonic output also if the screen-grid be at a higher positive potential than the anode. The valve then acts as a dynatron, the anode circuit tending to oscillate at the harmonic frequency to which it is tuned and reinforce the harmonic output. Used in this way the circuit becomes a dynatron frequency-multiplier, crystal-controlled. The mechanism of frequency-multiplication will be better understood after we have studied amplification by valves, and will be treated in more detail in the chapter dealing with telephony. At the present juncture we will be content to note its possibility and to consider how it influences the production of radio-waves at constant frequency.

It has been shown that a few watts of oscillatory energy can be obtained from a master oscillator controlled either by rigid design and constant temperature or by a piezo-electric oscillator. The frequency of this unit cannot be extremely high and for best stability should be lower than that to be transmitted. It is possible by careful design to maintain this frequency constant at a selected value to within better than one part in 10,000, and in fact to make the maximum variation negligible for all practical purposes.

A third method of achieving the same result has been mentioned, in which a tuning-fork provides the basic frequency. This will be of a few thousand cycles only, and the fork will be electrically maintained and in a constant-temperature oven. Successive higher harmonics of the fork frequency are then selected by successive valve multipliers until the desired radio-frequency is reached. Many stages will clearly be needed even if each multiplies several-fold, and hence the equipment is elaborate. It has the advantage of excellent constancy, however, and is useful in the synchronizing of radio-stations which are to work on an identical frequency, because the actual fork frequency is low enough to be sent from one station to the other over cable, either for direct remote control or for comparison purposes. The system finds favour in the case of medium-wave-length commercial or broadcast stations, but is not often justified when very high frequencies are needed, the quartz crystal being very much simpler. One effective form of circuit whereby

very high harmonics can be selected at good amplitude is a type of push-pull valve oscillator termed the 'multi-vibrator', and this is much used in the laboratory comparison of frequencies against the harmonics of a standard fork. It will be described later in its application to television.

Modern low-power or mobile transmitters may employ a simple crystal-controlled oscillator coupled to an aerial circuit, the crystal being ground to the radiated frequency. This is quite practical at medium frequencies between some few hundred kilocycles and 10 megacycles, for powers of at most 100 watts, and for hand-speed telegraphic sending. It cannot yield the best frequency stability, however, particularly when telephony is desired, and is mainly useful when the equipment must be small and light. The normal procedure in commercial or higher-powered installations is to follow the master oscillator by a succession of frequency-multiplying and amplifying stages, whereby the frequency is raised and the energy increased until it is sufficient to excite the grids of a final bank of large power output valves feeding the aerial. There is no recognized or invariable sequence of stages, since these are determined by the final power and frequency needed, the views of the designer, and considerations of cost. For example, it is possible to raise the frequency to the final value in a series of low-powered stages, and follow these by a series of amplifiers until the power becomes sufficient. Alternatively, the amplitude can be raised in successive stages concurrently with the frequency in a series of successive multiplying and amplifying stages, a method favoured in high-power transmitters because it simplifies the isolation of stages. It is easy to amplify to a moderate extent in a single stage, but the use of several stages at the same frequency involves elaborate screening to prevent unavoidable reaction and risk of oscillation within the amplifiers. Changing the frequency at intervals overcomes this difficulty, since very little reaction can occur by stray couplings between circuits at a different frequency. Thus one system of design employs alternate multipliers which select the third or fourth harmonic of the input, followed by amplifiers to make up for the energy lost in multiplying and to give an over-all increase. In any case the system should terminate in a large power amplifier working at the frequency to be radiated.

Now in frequency-doubling it is possible at all but the highest

frequencies to obtain an output from modern valves which is also amplified and represents increased oscillatory potential and probably increased energy over the input to the grid. This opens up what is probably the most convenient method of design, that in which both multiplication and amplification are combined in each stage. Such a transmitter comprises a master oscillator followed by a series of valve stages, each of which delivers an amplified power output at double the frequency of that preceding it, finally driving a power amplifier in which no

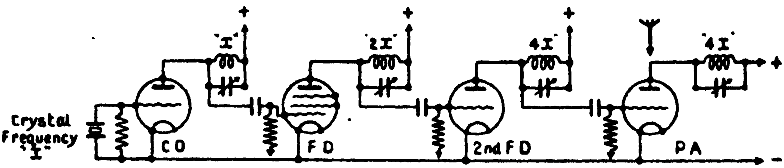


FIG. 65.

frequency change occurs. As an example, suppose we require an energy of 1 kilowatt to an aerial circuit at a frequency of 10 megacycles, which represents a medium-powered transmitter working at 30 metres wavelength. A four-stage design would be suitable, as shown in Fig. 65, the first stage being a crystal oscillator working at 2·5 megacycles, a frequency near to the optimum for best crystal efficiency. If a triode, this might deliver 5 watts of radio-frequency energy to the grid of the following stage, which might be a large pentode valve with advantage. The conversion efficiency of pentodes being high, we might expect from 20 to 50 watts of energy to be available at the first harmonic to which its anode circuit was tuned, namely 5 megacycles. This energy would supply the grid of the next stage, probably a large triode valve acting as doubler and providing from 50 to 100 watts of energy at 10 megacycles from its anode circuit. This would be at the desired frequency, and would provide ample energy to excite the grid of a still larger valve capable of delivering 1,000 watts to the aerial circuit. This transmitter would be described as a crystal oscillator followed by two frequency-doublers and a power amplifier, often expressed by the notation 'C.O.–F.D.–F.D.–P.A.', showing its main features in a convenient form.

Fig. 65 shows the arrangement in diagrammatic form; the circuit details will be treated in turn in the following chapters.

It offers several advantages over any form of self-excited oscillator. The frequency is determined by the crystal oscillator, and is not affected by the adjustment of the circuits which follow or by the aerial coupling. To ensure this, it is best that the stage following the crystal oscillator should operate with 'class A' bias, and draw no grid current that might damp the former stage. It is then termed a 'buffer amplifier'. The frequency can be fixed to any exact value, and maintained there with any desired accuracy, thus enabling distant receivers to be of the highest selectivity and permanently tuned, whilst other services can work with the minimum possible frequency-separation. There is no danger of the transmitter's drifting in frequency by more than a known amount, or getting onto the channel of another station through misadjustment. Moreover, anything which might happen to disturb the later stages will not affect the frequency. Valves can be replaced without risk, and the effects of weather upon the aerial, or swinging in the wind, will not affect it. Still more important is the fact that telegraphic keying or telephonic modulation can be introduced into one of the later stages, where they can have no effect upon frequency, a condition impossible in any of the earlier transmitting systems.

EXAMINATION QUESTIONS

1. Show by means of a diagram how a three-electrode valve can be used to generate oscillations. Show in the diagram means whereby the direct-current anode potentials and grid-bias potentials are excluded from the high-frequency circuits.

City and Guilds Institute. Grade 1, 1932.

2. Describe, with diagrams, how ultra-short waves may be generated and modulated for the broadcasting of speech. (See also Chapter IX.) *Institute of Wireless Technology. May 1935.*

3. Outline three methods by which the frequency stability of a continuous-wave wireless transmitter can be improved.

4. Explain how a quartz crystal can be used to determine the frequency radiated by a wireless transmitter. What are the principal advantages in practice from this system of wavelength control?

5. Explain the use of valves to multiply the frequency generated by an oscillator, in such a manner that a higher frequency can be radiated. What relation must exist between the original and produced frequency? What is the most effective condition of biasing for valves used in this manner?

6. A three-electrode-valve oscillator with the oscillatory circuit connected to the anode is required for a wavelength of 600 metres. Taking the resistance of the oscillatory circuit as 30 ohms and the internal resistance of the valve as 5,000 ohms, find what the values of the inductance and of the capacitance should be in order that the valve output may be a maximum. *Grad. I. E. E. 1935.*

7. Describe with a diagram the construction of one type of high-frequency alternator used for radio-transmission. What methods of keying are adopted with such machines?

C. and G. Intermediate 1935.

8. Give a diagram of a simple valve oscillator in which the high-tension and grid-bias potentials are excluded from oscillating circuits.

C. and G. Intermediate 1935.

9. Explain the action of the Duddell-Poulsen arc transmitter, stating what you know of its construction, uses, and limitations.

10. What is the effect termed 'negative resistance'? Show that this condition is favourable to the generation of oscillations, and mention the necessary conditions under which these can be maintained.

11. What changes must be made in the simplest form of valve oscillator circuit to enable it to generate short waves? Sketch four stages in the evolution of such a circuit, explaining the reasons for the modifications shown.

12. What is a 'linear oscillator'? Explain how a construction of this nature improves the performance and stability of an oscillator when used to generate ultra-high frequencies.

13. Name and briefly describe two methods whereby oscillations at wavelengths below one metre can be generated by valves. Why are reaction circuits useless for this purpose?

14. Describe the essential circuit and properties of the following types of master oscillator: (*a*) crystal-controlled valve, (*b*) electron-coupled, (*c*) Franklin or 'high-Q'. For what classes of use would you select each type?

CHAPTER VII

DETECTION AND RECTIFICATION

We have seen in previous chapters that signals sent out from the transmitting aerial are conveyed through space by trains of electromagnetic waves. We have now to consider how the signals transmitted by means of these waves can be received and made perceptible to our senses.

If a train of electromagnetic waves encounters a conductor, it will set up in it oscillatory E.M.F.s and currents of the same frequency as that of the waves. The first requirement of a receiver, then, is some form of electrical circuit which will respond readily to the effect of electric or magnetic forces in the waves. Such circuits may consist of an elevated aerial wire or array, or of a frame of wire tuned by a condenser, or even a wire laid on the ground. For the present, however, we will confine our attention to ordinary elevated aerials.

Since the oscillations set up in the aerial are of similar frequency to that of the waves, we next require some device for translating the electrical energies of the waves into a form which is perceptible to our senses. The device usually employed is a pair of telephones, or a loudspeaker, but a sensitive galvanometer or some arrangement for causing the signals to record themselves automatically on paper may be used.

Such instruments as these, however, cannot function at the high frequency of the received waves, and some further arrangement is therefore required for transforming the high-frequency currents into currents of a frequency to which the telephones or recording instrument can respond. Such devices are called *detectors*. Crystals and two- or multi-electrode thermionic valves are examples.

There are, of course, many circuit arrangements by which the detector and telephones can be connected to the aerial to form the simplest possible type of radio-receiver. One such is that shown in Fig. 66. The aerial is connected to earth by the coil L and condenser C_1, which is adjustable. The aerial is tuned by the adjustment of C_1 until its frequency is the same as that of the arriving wave. If the natural wavelength of the aerial is too large it can be reduced by the introduction of a series

condenser C_2. Thus if C is the capacity of the aerial system (made up of the capacities of the aerial and the condensers), and the self-induction of the aerial and the coil is L, then L and C are adjusted until

$$\lambda = 1885\sqrt{(LC)}$$

where L is measured in microhenries and C in microfarads and λ is the wavelength to be received expressed in metres.

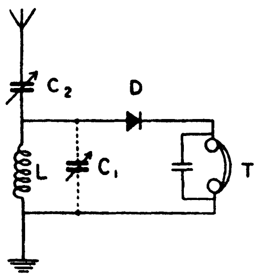


FIG. 66.

In the figure, D represents a crystal detector and T the telephones. The incoming waves set up oscillations in the aerial and cause a difference of potential to be created across the coil L. Thus high-frequency alternating potentials are applied across the detector D which, as we shall see later, cause low-frequency pulses of current to pass through the telephones T.

The headphones employed for many wireless purposes consist of a small horseshoe magnet fitted into the case of the instrument. The magnet is to some extent permanently magnetized and is wound with a large number of turns of fine wire. The windings of the two poles are connected in series, so that the received signal passes through both windings. The ear-piece, which is usually made of hard rubber or some similar material, is screwed on to the case. Supported round its edge by the case is a diaphragm of thin, soft iron which is brought very near to the pole pieces, but without touching them. The sensitivity of the telephones depends largely on the correct adjustment of the distance of the diaphragm from the pole pieces. In some makes this distance is about 0·003 inch.

In Fig. 67, A represents the case of the telephone, D the diaphragm, M the core of the electromagnet, B the windings, and E the ear-piece. Since the electromagnet is permanently magnetized, the diaphragm is strained slightly and pulled towards the pole pieces. When the currents due to the received signal pass through the windings, the attractive force of the magnet on the diaphragm changes. The motion of the diaphragm thus produced translates the electrical currents in the windings into audible sounds. Obviously, to obtain sufficient movement with the small currents in the receiving apparatus,

the moving parts must be made very light and the magnetic field of the magnet must be as large as possible. Hence the windings round the magnets must consist of a large number of turns. The amount of wire and the number of turns in the windings are usually indirectly expressed by stating the resistance of the windings in ohms, e.g. 120, 2,000, or 4,000 ohms. (In general, the higher the resistance the more sensitive are the

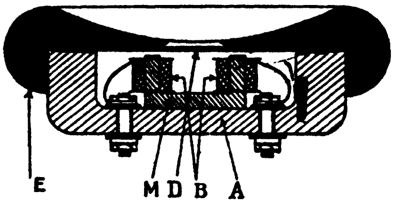


FIG. 67. Telephone Receiver.

telephones.) The practice has therefore grown up of speaking of telephone receivers as high- or low-resistance phones. It must be clearly understood, however, that the resistance itself of the telephones has no special virtue. The reason why a high-resistance telephone is more sensitive is that there are more turns of wire in the windings round the magnet, and therefore a feeble current acting in the winding will cause a larger magnetic force to act on the diaphragm.

The function of the detector is, as we have seen, to transform the high-frequency currents into currents of a frequency to which the telephones or recording instruments can respond. The frequencies of the electromagnetic waves used in radio-communications may vary from about 15,000 to many million cycles per second. A wavelength of 300 metres, for example, corresponds to a frequency of 1,000,000 cycles. Few telephone diaphragms, however, can respond to a frequency of more than 5,000 cycles per second; and the frequencies to which they vibrate most strongly are, as a rule, of the order of 1,000 cycles per second.

In the early damped-wave systems each dot or dash is made up of a number of groups of waves, or wave trains, following one another at a more or less definite rate. With a spark transmitter having a musical note the number of wave trains emitted per second may be between 100 and 1,000. In detecting such waves each group of oscillations is caused to produce a pulse of current

through the telephone, and the diaphragm therefore responds to the spark frequency and gives out the musical note corresponding to it. Continuous waves, as we shall see later, may also be reduced into groups of low-frequency oscillations which can be rectified by detectors so as to produce a musical sound in the telephones, and when this is done they are termed interrupted or modulated continuous waves, abbreviated to I.C.W. or M.C.W. An I.C.W. wave is actually broken up into sections, as by a commutator or an interrupted high-tension supply, whilst an M.C.W. wave varies in amplitude without actual interruption. Both are used to assist reception by simple equipment or to attract attention in emergencies. They are not essential for communication to-day, since plain continuous waves can be received by modern methods.

Many types of detectors have been used in radio-communications. Certain of these which have historical interest have been mentioned in Chapter I. At present, however, only two varieties are in common use, the crystal detector and the thermionic valve. Crystals and valves both detect by rectification, i.e. they turn the high-frequency oscillations into slow unidirectional currents.

Many different varieties of crystalline substances in various combinations are employed for this purpose. Examples of these are: a fragment of *zincite* (zinc oxide) in contact with a piece of *bornite* (sulphide of copper and iron) or *molybdenite*; or *galena* (lead sulphide) in contact with a metallic point of a fine wire; or *carborundum* (a compound of silicon and carbon) in contact with steel.

If different potentials are applied to any of the above crystal combinations, and if a graph is plotted between the value of the applied potential and of the corresponding value of the current across the point of contact, the result is what is known as the characteristic curve of the crystal. An example of such curves is given in Fig. 68, for zincite and bornite in combination. The curve shows that: (1) the resistance of the crystal combination does not obey Ohm's Law but depends on the magnitude of the current; (2) the crystal allows more current to flow through it when a given voltage is applied in one direction than when the same voltage is applied in the opposite direction; (3) at one particular voltage there is a more or less sharp bend.

If an alternating potential is applied to the crystal, the resis-

tance of the crystal will in general be different for the positive half-cycles of potential from its resistance for the negative half-cycles. The existence of the bend in the characteristic curve is the property which causes the crystal to serve as a detector or rectifier, and is characteristic of rectifying devices.

Suppose a train of oscillations is applied to a crystal combination having a characteristic curve similar to that shown in

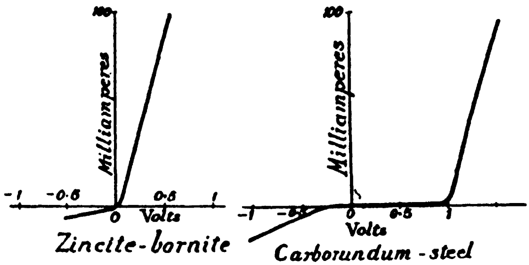


FIG. 68.

Fig. 68. In Fig. 69 the instantaneous values of the high-frequency potential applied are represented by the curve drawn below the characteristic. The maximum positive and negative values of the oscillating potentials are given by the points P_1, P_3, etc. and P_2, P_4, etc. respectively. If lines are drawn parallel to the current axis of coordinates, then the points in which these lines meet the characteristic curve represent the values of the current passing through the point of contact of the crystals for the values of the applied voltage corresponding to $P_1, P_2, P_3, P_4, \ldots$.

The instantaneous values of the currents are represented by the curve drawn to the right of the figure, the currents corresponding to the positive and negative maxima being represented respectively by the points C_1, C_3, etc. and the points C_2, $C_4, \ldots$, etc. It is at once seen that, owing to the shape of the characteristic curve, the positive currents are much greater than the negative currents, and that, taking the mean of the currents, the net effect is to allow a slow pulse of positive current, represented by the dotted line *a a a* in the figure, to flow through the telephones.

In the case of the carborundum curve and some others the position of the bend in the curve occurs not at the zero potential but at a point corresponding to a certain positive applied

potential. Thus, in order that the carborundum crystal may detect efficiently, the potentials due to the signal must be applied so as to cause oscillations to take place on either side of the point corresponding to this positive potential.

To bring about this result, the necessary potential is applied across the crystal by an arrangement called a *potentiometer*. In

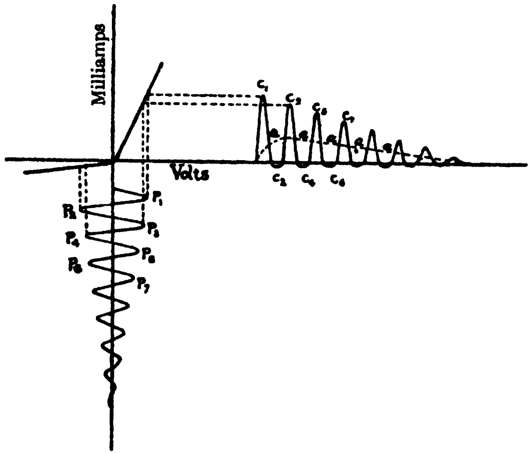


FIG. 69.

this instrument (see Fig. 70 (*a*)) the two terminals of a battery B are connected to the ends of a high-resistance wire P_1P_2. The fall of potential between P_1, P_2 is that of a battery of, say, 4 volts. If a connexion S_1 is arranged to slide along the resistance wire, any difference of potential between 0 and 4 volts can be applied between conductors A_1, B_1 attached to P_1 and S_1. For example, if S_1 is in the middle of P_1P_2, the potential between such conductors will be 2 volts. If it is required to apply either positive or negative potentials between the conductors, one is attached to S_1 and the other to the middle of the battery, as shown in Fig. 70 (*b*). Then, if S_1 is in the middle, the potential between A_1 and B_1 is zero, while the potentials of opposite sign can be applied according to whether S_1 is moved towards P_1 or P_2.

A potentiometer may be introduced in series with the telephones of any crystal receiving circuit, when it enables the working-point to be adjusted to the sharpest bend of the characteristic curve. In the days when the crystal detector was very widely used, a potentiometer would seldom be dispensed

with, but to-day, it is not often thought worth while to employ one. The working of a potentiometer has been described fully because it also forms part of many valve circuits and is useful in laboratory work. In the former case it is an excellent means of providing an adjustable grid bias of the precise value needed, or for varying this continuously when taking the characteristic curves of valves.

It can be shown that the rectified current due to a crystal depends on the square of the signal E.M.F. applied, as well as

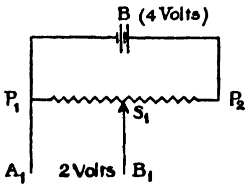


FIG. 70 (*a*).

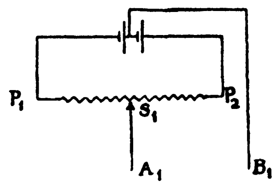


FIG. 70 (*b*).

on the steepness of the slope of the characteristic curve. Thus, if the signal strength is halved, the rectified current is only one-quarter of its previous value. Accordingly, the sensitivity of a crystal detector decreases rapidly for weak signals, and is termed a 'square-law' rectifying-device.

Although the mean current through the crystal is represented by the dotted curve in Fig. 69, nevertheless this mean current is actually made up of rapid high-frequency pulses. The telephone windings have a large coefficient of self-induction, and would, if they had free play, impose a sufficiently large reactance to choke out the rapid oscillations. The telephones are therefore shunted by a condenser. The rectified pulses charge up this condenser, which, when the wave train ceases, discharges the electricity through the telephones. Frequently the self-capacity of the telephone windings and the leads acts as a condenser of sufficiently high capacity to enable a special telephone condenser to be omitted; but, for satisfactory working, capacity in some form or other must be present across the telephones.

The physical reasons which cause crystalline materials to exhibit their sharply curved characteristics are not fully understood. It is probable that thermoelectric effects may play an important part, but such effects do not provide an explanation of all the phenomena observed. It may be remarked that the application of heat to some crystals entirely destroys their

rectifying properties. Not all specimens of the same material have detecting properties, and before a sensitive piece of crystal is found many others have to be examined and rejected.

The two crystals, or the crystal and metal contact, of a crystal detector are usually mounted in metal stands on an insulating base. The crystals are soldered in the metallic cups by a solder with a low melting-point. The surface of the crystal is not sensitive as a whole, and suitable points have to be found upon it by trial. The fine-wire contact or 'cat's-whisker' in a galena combination, or the piece of zincite in a zincite-bornite couple, is therefore carried on the end of a movable arm, so that it can be brought into contact with the other crystal at any desired point.

An arrangement is also usually provided for adjusting the pressure at the point of contact. As a rule, the pressure between the elements of the combination should be slight. Carborundum is unusual in giving better results in contact with a metal surface rather than a point, and in giving good results under considerable pressure.

Considerable space has been devoted to the crystal detector because of the admirable introduction it provides to the ideas of detection and rectification. As a simple detector the crystal is quite efficient, and its poor sensitivity can be overcome by the addition of valve amplification. When this is done, however, the stronger signals soon upset the delicate crystal adjustment, and so the combination of crystal and valves has gone out of use. The crystal still remains in use on a surprisingly large scale for cheap local broadcast reception, but in common with all other early detectors it has been entirely superseded by the valve in more advanced receivers.

At the beginning of Chapter V we saw how the two-electrode or diode valve is used as a detector, and that it was first developed solely for that purpose. The similarity of the characteristic curve and circuit of Fig. 34 to that of Fig. 68 will be obvious, and we will consider the reasons for the wide recent revival of the diode a little later.

The three-electrode valve can be used as a detector of wireless signals in two ways. An examination of the anode-current grid-volts characteristic shows that the form of the curve is practically the same as a crystal characteristic. If the high-frequency oscillations from the aerial are applied to the grid of a valve, and

the grid potential is adjusted so that before the arrival of a signal the anode current is that corresponding to the point P_1 at the bend of the anode-current grid-volts characteristic, then the incoming signal will cause the grid potential to oscillate on either side of this point. Let the values of the potential v_g

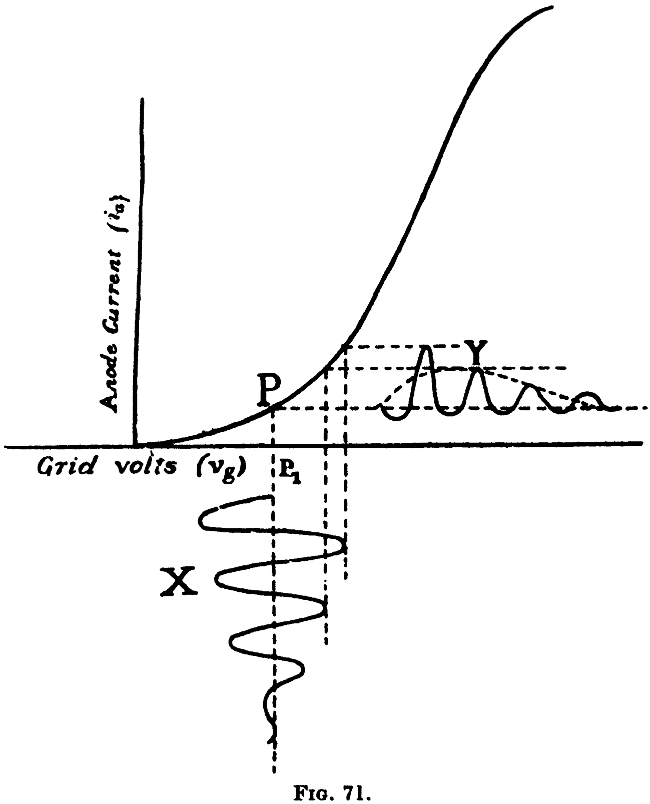


FIG. 71.

applied by the signal to the grid from moment to moment be represented by the curve X in Fig. 71; then the corresponding values of the anode current i_a passing through the valve to the telephones will be given by the curve Y. It will be seen that when the signal is causing changes of potential on the grid the anode current is oscillating on either side of its value at P_1. Owing, however, to the change of curvature of the characteristic at this point, the change in amplitude of the anode current for the first half-oscillation is greater than that for the second half-

oscillation, and hence, as in the case of a crystal detector or a two-electrode valve, the change in anode current due to the first half-oscillation is not completely neutralized by the second half-oscillation, and so on. Thus the train of oscillations produces a slow pulse of unidirectional current indicated by the dotted line drawn through the curve Y in Fig. 71. A series of wave trains similar to X will produce a series of unidirectional pulses

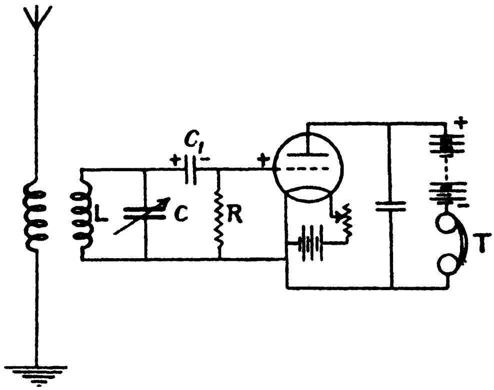


FIG. 72.

to which the telephones will be able to respond. This method of detection is known as *anode rectification.*

The detecting action is greater in proportion for strong signals than for weak signals (since with strong signals the successive half-oscillations of anode current are much more unequal) and thus again resembles the crystal.

The second and more sensitive method of using the three-electrode valve as a detector depends upon the form of the *grid-current grid-volts* characteristic. The circuit arrangement employed is shown in Fig. 72. The secondary circuit LC of the receiver is isolated from the grid by a small condenser C_1, and between the condenser and the filament is connected a high resistance R of the order of 1 megohm. (Instead of connecting this resistance as shown, it is sometimes connected as a shunt across the condenser C_1.)

Before the arrival of a signal the potential of the grid will be approximately zero. When the aerial is excited, however, high-frequency currents will be produced in the circuit LC. These currents will induce a charge on the left-hand plates of

the condenser C_1, which by the ordinary laws of electrostatics will attract a charge of opposite sign on the right-hand plates of C_1 and repel a charge of like sign to the grid. Thus, if the signal current is flowing in such a way as to charge the left-hand side of the condenser positively, the grid will also be charged positively, and will consequently have a positive potential. In

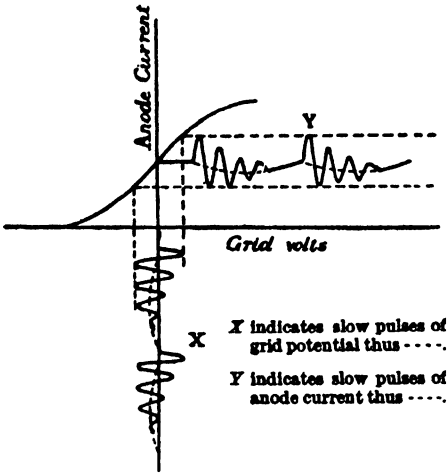


FIG. 73.

the next half-oscillation of signal current the grid potential will reverse and become negative. Thus during the reception of a train of oscillations the potential of the grid will oscillate on either side of zero potential. Now, as shown by the *grid-current grid-volts* characteristic curve (Fig. 37), negative electrons flow *into* the grid during the periods in which the grid potential is positive; and there will thus be an accumulation of negative electricity on the grid. Since the grid is isolated by the condenser C_1, the negative charge collected will remain on the grid and gradually make its potential more negative. Thus the actual grid potentials during the arrival of a train of oscillations will be as shown at *X* in Fig. 73. The effect of the oscillations of grid potential will be to cause changes of anode current, but since the grid is all the time collecting a negative charge the arrival of the signal will result in a slow net decrease of anode current, as shown by the dotted line of curve *Y* of Fig. 73. If the grid were completely isolated, the collection of negative

electricity on the grid would go on until the anode current fell to zero and the action of the valve was stopped. The presence of the high resistance between the grid and the filament, however, provides a path by which the negative charge of the grid can leak slowly away, so that the grid returns to zero potential. Thus the effect is to produce slow variations of anode current which will be audible in the telephones.

If the two methods of using the valve as a detector are compared, it will be seen that in the second method the anode current decreases to a minimum value and then returns to its mean value during the arrival of a train of oscillations. In the first method, however, when the lower bend in the anode characteristic is used, the anode current increases slowly from its mean value. Thus, since the changes of anode current affecting the telephones are in opposite directions in the two methods of detection, it is necessary, when the valve is adjusted for the second method, that the portion of the anode-current curve corresponding to the grid potential at which a grid current begins to flow should be straight and not near the bend in the curve. Otherwise a telephone current due to anode rectification will cancel to some extent the telephone current due to grid rectification, and the valve will thereby be less efficient as a rectifier.

These two methods of employing a three-electrode valve as a detector are generally termed anode rectification and grid rectification, or frequently more fully anode-bend and leaky-grid rectification respectively. A more recent modification of the latter is termed power-grid rectification, implying that the valve is worked under conditions such that it can handle appreciable power, as contrasted with ordinary detection, in which we are concerned essentially with the conversion of alternating into unidirectional potentials. This is done by the use of high values of anode voltage and somewhat greater anode current, whilst the grid leak is reduced from the conventional value of from 1 to 10 megohms to 250,000 ohms or even less. The reason for this is to enable the valve to rectify larger potentials without overloading, and the need for it will soon become apparent. Multi-electrode valves such as the screen-grid or pentode can be used as detectors but do not differ from the triode in the essentials of their operation. They offer certain advantages due to their higher amplification factor—which

results in larger changes of anode potential for a given input signal—and to secondary circuit conveniences which will be understood when we come to consider complete receivers.

The simple methods of considering detection that have been given are useful in the understanding of the physical action of these circuits and are adequate when the reception of simple telegraphic signals only is under discussion. In the modern receiver, however, there are two additional factors which must be taken into account. One of these is the rectification without undue distortion of continuously modulated telephony signals, and the other the ability to rectify also without increased distortion the very strong signals which can result when the detector is preceded by amplifying valves. To understand how these requirements are met we must study the general problem of rectification from a somewhat different angle, and must also assume in advance some knowledge of telephonic modulation contained in Chapter IX.

Telephony is transmitted by a continuous carrier wave. This is not broken up into the long and short periods of the Morse code, and is not tone-modulated or broken into trains of damped waves as in the various telegraphic systems, but is of absolutely constant amplitude and frequency at all times when no telephone signals are being transmitted. When speech or music takes place the amplitude of this wave is varied in exact accordance with the microphone currents. We know that in the telephone a continuous direct current varies in amplitude to correspond to the air-pressure waves produced by sounds reaching the microphone. To transmit this over radio the amplitude of the carrier wave is varied so that the envelope of the high-frequency oscillations corresponds to the variations of telephone current. The state of affairs is illustrated in Fig. 74, where (*a*) represents the variation of microphone current (or potential) produced by the sound, and (*b*) the form of the resulting modulated radio-wave. To convert this into a form capable of actuating a telephone ear-piece or loud speaker we have seen that it must be rectified, and for the current through this telephone to be an exact copy of the original microphone current, as is essential for faithful reproduction, the rectification must be such as to remove all that part of the signal below the zero line. The result is a potential varying as in (*c*). This is unidirectional and can operate the telephones; but, as we have

seen, only the envelope of the high-frequency pulsations will produce a sound, and this will be an exact copy of the original if the telephones themselves are free from distortion.

Now, it can be shown that the modulated wave of Fig. 74 (*b*) is equivalent to the product of the high-frequency carrier wave and the low-frequency sound waves. This could be confirmed by multiplying these two wave forms graphically, when the product would be identical with (*b*), and incidentally would show up the important fact that the actual frequency of the modulated wave is not constantly that of the carrier but is altered by the modulation. The meaning of this will be examined later when we discuss selectivity. These high-frequency and low-frequency components cannot be easily separated, and rectification by a method which completely suppresses all the negative half-waves as in (*c*) converts the wave into a form which is the *sum* of the original high-frequency and low-frequency components together with a direct-current component. The curve of (*c*) is equivalent to the sum of these three, and so it becomes clear that the rectifier cannot in itself separate the modulation from the carrier wave. It merely converts the product of the two into the sum, and their separation is effected by the circuits which follow. Some form of filter must follow the rectifier, able to separate the modulation from the high-frequency pulsations, and without this a telephony modulated signal cannot be made audible.

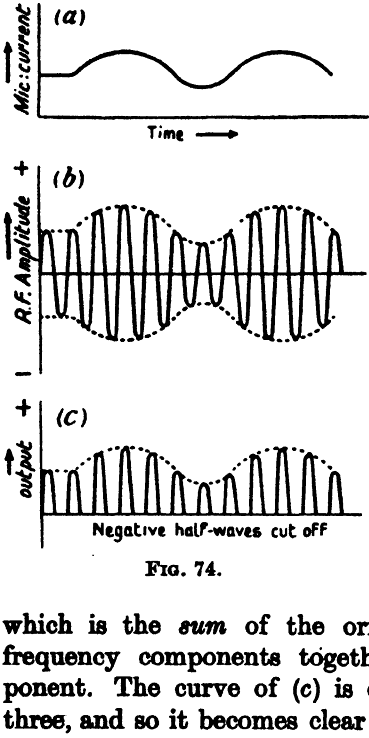


FIG. 74.

It is now easy to understand the true function of the condenser connected across the telephones in the simple crystal circuit of Fig. 66. This forms a simple filter, depending for its action upon relative reactance. The condenser offers a very low reactance to the high-pulsation frequency, and hence the bulk of this component flows through this rather than through the

telephones, which are inductive and offer a very high reactance to high frequencies. On the other hand, the lower modulation frequencies are inversely treated and find a lower reactance path through the telephones than through the condenser. This is the simplest form of filter and must follow any rectifier if modulation is to be reproduced. It is illustrated in Fig. 75. Here the mixed components after rectification pass through two parallel paths, the one consisting of a condenser, through which the major portion of the high-frequency pulsations pass, and the other of a resistance, inductance, or both, through which only the lower-frequency modulation currents can flow. Such a simple filter is not entirely effective, since a small proportion of the wrong component must pass through each path, the reactance of which is not infinite; but it is quite effective enough for most practical purposes. We shall see later cases when a more complete separation is necessary. It should also be noticed that some slight distortion of the modulation frequencies is inherent in the separation, because the condenser will have a lower reactance to the highest of these than to the lowest, and hence a greater proportion of the higher modulation frequencies will be wasted by leakage through the condenser. This weakens the highest tones in speech and music somewhat, and must be taken into account when aiming at the best possible reproduction.

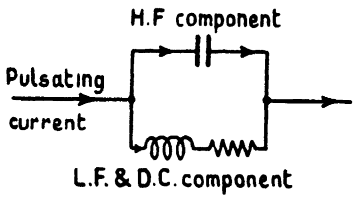


FIG. 75.

An alternative point of view may help toward the understanding of the above process. Most devices, such as the telephones, cannot respond at the high speed corresponding to the rectified carrier wave. Even after these have been rectified and rendered unidirectional they could not therefore be heard. A condenser is able to store up, or integrate, these brief impulses, so that the potential across it varies as the envelope of the curve of Fig. 74 (*c*). This slowly varying potential can drive a current through the telephones.

It is often incorrectly thought that mere rectification will make signals audible, but this will not be the case unless a condenser or integrating circuit is also present to set up the necessary low-frequency potentials.

Consider now the merits of the various detectors we have described in the light of this extended theory.

The original diode is clearly the most perfect, possessing no conductivity whatever in one direction. It will suppress the negative half-waves completely, giving an output exactly as in Fig. 74 (*c*). Moreover the diode is unaffected by the amplitude of the signals and very little by the depth of modulation, so that provided that the maximum cathode emission is never exceeded there can be no distortion within the diode itself. This fact explains the increasing use of the diode in modern receiver design.

The crystal and the anode detector exhibit a less perfect curve in which there is some conductivity in the reverse direction and which will not give complete rectification of weak signal potentials. Moreover the rectification has been shown to become better when the input is greater, and hence distortion must occur. This will be of the type known as 'amplitude distortion', since high-amplitude modulation will be better rectified and better reproduced than that of lower amplitude. This effect is to be expected, since the rectification of both anode and crystal are square-law for moderate amplitudes, whilst that of the diode is linear. To overcome distortion in the anode-bend rectifier it is necessary to amplify the signals before detection so that only large oscillations are applied to the valve grid. These will swing the operating-point well on to the straight portions of the characteristic curve, to the right of P in Fig. 71, and in this region distortion is minimized and rectification approximately linear. Little advantage remains from the use of anode rectification after this is done, however, since the sensitivity is little better than the diode and distortion always somewhat greater; and with the almost unlimited amplification given by modern valves the diode has become preferable.

To understand the distortion produced by a grid rectifier it is best to treat the valve from a somewhat different angle. The grid and cathode of this circuit can be regarded as forming a diode combination. The incoming signals are applied between these two electrodes, and are therefore diode-rectified without distortion, producing, upon the grid, potentials similar to those in Fig. 74 (*c*). The remainder of the valve can now be regarded as a low-frequency amplifier in which these grid-potential variations are amplified to form

the anode current. Since both the high-frequency and low-frequency components of the rectification process are amplified, the anode current represents an amplified version of (c) and separation by filter is still needed. Unfortunately, the working-point on the characteristic at which the leaky-grid detector operates will be in the neighbourhood of the bottom bend of the curve, instead of at the mid-point of the straight portion, which we shall see is the best for distortionless amplification. Should the signals be very strong there is a tendency for the grid to bias itself strongly negative, when the operating-point may reach or even pass below the bottom bend of the curve, and part of the modulation wave form may be cut off in the amplification process. Thus it is clear that the grid detector must introduce some distortion which increases at high amplitudes, and that the action of the circuit can be regarded as a combination of distortionless diode rectification with imperfect amplification in the same valve. In view of these facts the grid detector has lost the popularity which it possessed a few years ago, giving way to the equivalent but more perfect combination of a diode rectifier followed by a correctly biased and distortionless amplifying stage. We have seen when discussing new valve types that these two functions may be combined into one valve by the provision of diode and triode elements operating from a single cathode. This is the diode triode or similar dual valve, and offers the same amplification and sensitivity as either the grid or anode detectors, with less liability to distortion.

The grid detector can be improved to a considerable extent if steps be taken to maintain the operating-point near the middle of the straight portion of the characteristic curve, and to lengthen this straight portion so that strong signals can be safely amplified. This is done in power-grid detection, in which high anode voltages are used so that the curve becomes long, and a low grid-leak value prevents the grid from charging up permanently to an excessively high negative potential. For several years this form of detection was the commercial favourite, but it has now given place to the somewhat more perfect diode whenever high quality of reproduction is desired. It is still a very widely used system for short-wave reception and whenever high sensitivity is a major consideration without the use of very many amplifying valves.

High sensitivity arises from two causes, one of which is the

amplification given by the triode aspect of the valve and the other the fact that a degree of rectification occurs from diode action at the grid no matter how weak the incoming signal may be. The former of these properties makes it possible to couple the anode circuit back to the grid to provide reaction, a portion of the amplified high-frequency component being utilized to intensify the original grid potentials, but not to a sufficient extent to produce continuous oscillation. In this way it might be considered that a portion of the detected signal is reamplified and redetected indefinitely, and a large increase in the final output from a given weak input signal can be built up.

Another way of regarding amplification by a reacting detector is to notice that the damping of the tuned circuit connected to the grid is reduced by its use in the manner we discussed when studying valve oscillators. A given amount of signal energy will build up a larger oscillatory potential across this circuit if the damping be reduced by reaction, and hence will set up larger grid potentials. By the critical adjustment of reaction to a value just insufficient to set up continuous oscillation, the combined effect of this reduced damping and progressive re-amplification can produce enormous sensitivity to the weakest signals, equal to that obtainable from a diode after several stages of amplification. It thus provides a highly sensitive receiver with a minimum of valves, and is of the utmost importance in economical reception. It is also invaluable at very high frequencies, where high amplification from valve stages is difficult to obtain by other methods.

When we come to discuss receiving circuits we will examine the conditions for best filtering and coupling the anode circuit of detectors to the following valves. There is one form of detection, however, in which these processes are assisted by the rectifier itself, and which is very useful at short wavelengths, where the more usual circuits tend to become less satisfactory. This is push-pull detection, shown in Fig. 76. It is not important what type of rectifiers are used, all those we have examined being equally applicable, but in the diagram are shown a pair of grid rectifiers connected in the push-pull manner. Their respective grids receive signal potentials from opposite ends of the tuned circuit *LC*, and therefore in opposite phase. This means that, whilst the anode current of one valve is at its

maximum, that of the other will be a minimum. Clearly a high-frequency pulse which is positive with respect to V_1 will be negative with respect to V_2, and vice versa, and so the two valves each rectify half of the incoming waves. All the half-waves will be rectified by the two valves together, and for this reason the rectification is said to be full-wave, and will of course make better use of the available signal energy. The respective anode circuits will each receive rectified pulses in the same direction

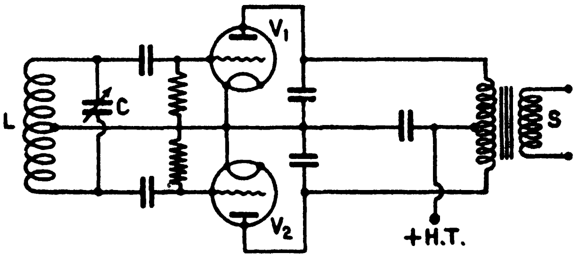


FIG. 76.

but alternate in time, and if connected as shown these will combine in the output to give twice the energy that would be obtainable from one valve. Full-wave rectification is clearly more efficient and somewhat more sensitive, but this is not the only reason why the circuit is of value.

It will be noticed that the valve grid-cathode capacities are in series across the input tuned circuit. They will therefore place less capacity and load across this circuit, which is advantageous at short wavelengths in receiving circuits just as it was in transmitting circuits. It allows a greater inductance to be used, and this results in a larger high-frequency potential being built up. It is true that only half of this is applied to each grid, but this loss is not nearly so large as might be imagined, because a larger total potential is to be expected from the tuned circuit under conditions of reduced damping.

The form of anode circuit shown employs two condensers to filter away the high-frequency pulses, as in single-valve detection. The tapped-primary transformer used to separate the modulation or lower frequencies from the high-frequency component and to transfer them to the following valve grid forms a further advantage of the system. By suitable design the secondary winding is placed symmetrically with respect to the

two sections of the primary, and thus stray capacity couplings which might allow high-frequency potentials to reach the following stage can be reduced. The potential of the secondary winding as a whole can be made more nearly equal to that of the cathodes. The low-frequency pulses add in the secondary winding and are completely transferred to the later stages. A characteristic of this circuit arrangement is that even harmonics which may be introduced into the original wave form by imperfect rectification are cancelled out by the differential action of the split anode circuit and do not appear in the secondary winding. This reduces certain types of distortion considerably and is often very helpful in receiver design.

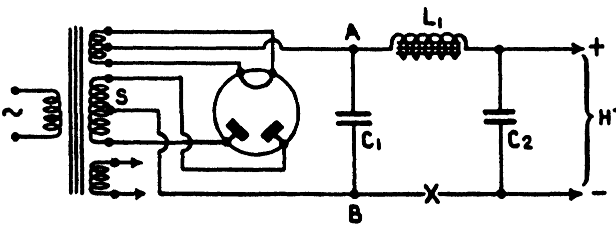


FIG. 77.

We have now considered all the rectifying systems in common use to-day for the detection of radio-signals. There is another use of rectification which has become very important with the increasing use of electric supply mains and the practice of operating receivers directly from them in place of batteries. This is the rectification of alternating current to produce direct current for valve anode supply. It is carried out by methods exactly similar to those described, but adapted to provide larger currents and potentials, and to work efficiently at low frequencies of from 25 to 100 cycles per second. By far the most general is the full-wave circuit employing two diode elements in what amounts to push-pull connexion, shown in the circuit diagram of Fig. 77. It is almost identical with the last circuit described, for the push-pull rectification of radio-signals, but with very different values of components. The alternating-current mains are taken to the primary of a mains transformer, which may be tapped at several different inductance values so that it can be used for differing mains voltages. By the selection of suitable primary turns it is possible to maintain the

secondary voltages always near to their nominal value, and to keep the primary inductance just sufficient to prevent excessive current from following. The transformer is provided with at least three secondary windings, one of which delivers the current to be rectified at a voltage generally higher than the mains, a second winding giving a low voltage which heats the rectifying valve cathodes, and a third or additional windings which supply heater current to the other valves in the receiver.

The high-voltage secondary S corresponds to the tuned circuit which supplied the signals to the push-pull detector valves of Fig. 76. The operation is similar, and a pulsating unidirectional current is produced between the points A and B which can be regarded once again as the sum of the mains frequency and a direct-current component. It is now necessary to separate the direct current from this pulsating component, which would otherwise cause hum in the receiver, and this is done by a capacity-resistance filter of the type we have been considering. A large condenser of several microfarads capacity, C_1, provides a low-reactance path through which the pulsating current will flow. The resistance is replaced in this case by a choke having high inductance and comparatively low resistance, the former being perhaps from 20 to 100 henries and the latter from 25 to 500 ohms. To get this ratio of inductance to resistance the choke must be provided with a laminated iron core. The choke offers a high reactance to the pulsating currents but a low-resistance path to the desired direct current. To remove traces of pulsating current which may pass the choke, and also so that the whole system may represent a low impedance as viewed from the receiver, a second large condenser C_2 follows the choke and completes the smoothing action. The whole smoothing circuit then becomes an example of a single-section π-type low-pass filter, and can be treated in that light when convenient. Occasionally the direct-current output is found to contain a small pulsating component which is more than can be tolerated, and further smoothing is introduced by the addition of a second choke and condenser, converting the filter to a two-section π network. Alternatively, it is sometimes an advantage to add a second choke at the point marked X in the diagram, thus raising the total reactance which impedes the pulsating component and making the filter symmetrical with respect to earth.

An arrangement of this kind whereby high-tension and low-tension current for a receiver is provided from the mains is termed a battery-eliminator. When first introduced the circuit was frequently made up into a self-contained unit which could actually replace existing batteries, and which would then contain also a potential-dividing network of resistances whereby various voltages might be tapped off to suit the various stages of receivers. More recently it has become increasingly usual to build receivers for mains operation exclusively, when these

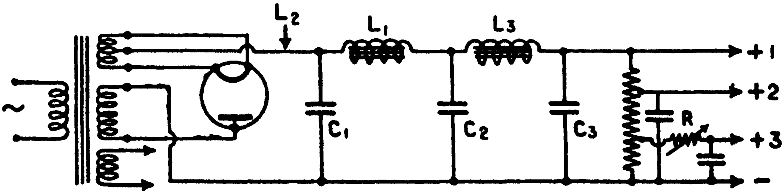


FIG. 78.

resistances are combined with the receiving circuit and form part of the anode circuit of each valve. Fig. 78 illustrates the smoothing and potentiometer network of an eliminator employing a two-section filter and providing for a separate voltage supply to three different valves, one of which supplies can be adjusted by the resistance *R*. It also illustrates the half-wave form of rectifying circuit employing a single diode valve only. This circuit is very similar in practice to the full-wave type, but is sometimes preferred for reasons of cost. It dispenses with one valve and employs a smaller transformer, but can only provide half the total current from a given size of valve. Also, since the pulsations are now at mains frequency instead of double this, condensers of twice the capacity are needed to give an equivalent degree of smoothing. For these reasons the circuit is used when only small output is required, and at a low voltage (of some 200) enables large-capacity condensers to be used cheaply. At higher voltages there is every reason to prefer the full-wave circuit, which also becomes the most economical when the current output exceeds some 60 milliamperes. It is interesting to notice that the full-wave circuit can be regarded as two half-wave circuits combined and employing a centre-tapped secondary in place of two separate transformers; it will be easily understood when considered in that way. The use of a centre-tapped winding is

helpful because the direct current flowing through this will be opposite in direction through each half, and will therefore produce equal and opposite magnetization in the iron core. There is no additional magnetization from this cause, and the core need only be of large enough section to carry the alternating magnetic field without saturation. This reduces the size of the transformer and makes it cheaper than the half-wave type, in which the core must be increased to withstand the direct-current magnetic flux. It leads to a considerable economy in the larger transformer sizes.

A good battery should possess very little internal resistance, and its voltage should vary very little with varying load. This property is termed voltage regulation, and any source of electrical energy is said to have good regulation if its potential varies little whatever current be drawn from it, whereas if the regulation be poor there will be a marked fall in potential with increasing current. It will be clear that whether the bad regulation be due to high internal resistance or not it can be expressed as an equivalent resistance. It could also be expressed as the fall in potential per unit increase in current. In many cases neither of these is satisfactory, since the rate of potential change varies with the actual value of the current and may not be a linear function, when the most satisfactory method is to show the relationship in the form of a curve.

Many receivers will not work satisfactorily unless the regulation of their high-tension supply be good, and for an eliminator to be satisfactory it should possess regulation comparable to that of a battery. Mains-operated receivers are sometimes designed to tolerate poorer regulation, but there are other circuits which we shall see later need very excellent regulation indeed for best results, and it will be useful to know how this property can be improved in eliminator circuits. The widely used circuit of Fig. 77 is not perfect in this respect, as the components clearly possess resistance. The major factors which cause the output voltage to fall with increasing current are: (*a*) The resistance of the transformer windings and other transformer losses. These can be summed into a factor termed the transformer regulation. This should be very small for a well designed transformer working within its rated load, but increases very rapidly if the design be skimped or the load exceeded. (*b*) The internal resistance of the rectifying valves. This may

be quite appreciable if valves be high-vacuum types, but is negligible for mercury-vapour-filled rectifiers. (*c*) The resistance of the smoothing chokes. (*d*) A factor depending upon the efficiency of rectification, and which is improved by the use of high capacities at C_1.

For best regulation the choke resistance (*c*) is a predominant factor and must be small. To retain ample smoothing with a choke of low resistance and somewhat limited inductance necessitates large condensers, and this is also desirable from the effects of (*d*). Hence modern practice is to use up to 8 microfarads for C_1 in place of the 1 or 2 microfarads thought sufficient in early designs, except at very high voltages, where the cost of such large condensers may be prohibitive. C_2 cannot be too large, and 8 microfarads is also a typical figure. Where additional smoothing is needed it will be best for regulation to obtain this by large condensers rather than by added chokes, which must increase the internal resistance. Attention to these points will result in good enough regulation for normal receiving purposes, but when still better results are necessary it must be supplemented by the use of mercury-vapour valves and a circuit modification shown in Fig. 78. This consists in the use of a choke directly after the rectifier, as at L_2. The filter then becomes one of inductive input in place of capacitive input, and it has been found that a choke of low resistance and a few henries inductance at this point can greatly improve the regulation, but at the expense of reduced output voltage. These alterations enable the regulation curve to be made nearly flat up to a certain maximum load, and give results comparable to the best battery supplies. If the first type of circuit is used, however, it gives higher output, and the curve will fall slowly in that case up to the maximum working load, after which the fall becomes increasingly rapid.

The potential given by an eliminator is of the order of the transformer secondary voltage (R.M.S.), or somewhat greater on light load, but its calculation involves too many factors to be given here. In addition to all the above sources of resistance, it is affected by the value of C_1 and of L_2 if used. If, however, the effect of these in conjunction with a given valve is known from the valve-makers and all other factors are known, the calculation of voltage for any particular current involves only successive simple applications of Ohm's Law. In the same way regulation varies too widely for any figures to be given, but is so nearly

linear at low currents that it can be approximately estimated if the eliminator be treated as a source of constant potential in series with its own total internal resistance, obtained by adding that of all the components. Recently methods of obtaining virtually perfect regulation have been developed which will maintain the voltage steady to within fractions of a volt. Some of these involve valve circuits designed to draw a current varying inversely with potential, and are thus not dissimilar to a negative-resistance device which compensates for the positive resistance of the eliminator or acts in the manner of a 'floating' secondary battery across the supply. Another is the 'stabilivolt', a device of the neon-tube type having characteristics such that the potential between two electrodes immersed in an ionized discharge is very nearly independent of the current drawn from them. These methods are not usually justified but can be resorted to in special cases such as equipment designed to operate consistently without manual adjustment, or in amplifying systems for television.

Eliminator circuits for transmission differ in no essential features from those used in reception or lower-powered equipment. Identical circuits employ larger voltages and currents and components suited to these. Rectifying valves can be used in parallel banks to increase the available current, and the problem of smoothing large currents is assisted by the fact that the power stages of transmitters do not demand such perfectly uniform current as do sensitive receiving circuits. Early transmitters were frequently supplied from large banks of secondary batteries and from direct-current generators producing the correct potential used on the valves. These methods have largely disappeared except for filament-heating and grid bias, and the high tension for high-power stages is derived from rectifying and smoothing equipment of the types described. The problem is often simplified by the use of alternators intended for this sole purpose, which produce alternating currents of higher frequency, such as 500 cycles per second. The pulsations resulting from rectification of these higher frequencies are much more easily smoothed by condensers of lower capacities since the reactance of a given capacity falls inversely as the applied frequency. The use of three or more phase supplies is also a great assistance, since the multi-phase pulsations which result are multiplied in frequency and reduced in amplitude by

the number of phases used, again simplifying the filtering problem.

Eliminator circuits for use on direct-current mains of course need no rectifying valves or transformers, since the direct current can be applied at once to a smoothing filter and potential-dividing network, with perhaps a series resistance to limit the current in case of accidents. It is usually found that the ripple on direct-current supplies is sufficient to necessitate smoothing,

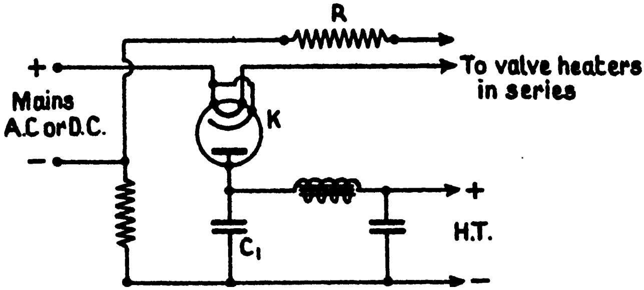


FIG. 79.

and frequently this must be very complete. Of interest is the circuit now used in the production of eliminators which will work equally well from either alternating-current or direct-current mains, and which has made possible the 'universal' receiver. It is shown in Fig. 79. Here no mains transformer is used and the mains voltage is applied directly to the rectifier, which must be half-wave. No step-up of voltage therefore takes place, and the high-tension output will be of the same order as the mains voltage. When the mains are direct-current, the rectifier simply acts as a resistance, allowing current to flow in the correct direction only through the path provided by its cathode emission. The heating current for all valves is derived directly from the mains also, and this was very difficult to achieve with earlier valves taking at least one ampere of heater current. If the current were limited to this value by resistance the power drawn from the mains at 230 volts would be 230 watts for heating alone, and with the high-tension current added to this the receiver would consume a quarter of a kilowatt per hour. Moreover the valve-heaters would not work reliably in series, whilst the use of filament types would introduce hum and biasing difficulties. The problem has been solved by the production of

specially designed 'universal' valves. These are indirectly heated, and the cathode-to-heater insulation is very high and will withstand full mains voltages without leakage which might set up hum. The heater is designed to work at higher voltage than usual, at correspondingly reduced current, and the latter is identical for all valves so that series operation of heaters is possible. Typical values are 0·25 ampere at about 16 volts per valve, in which case the total consumption from the mains by the heaters is reduced to a practical figure of 60 watts. Up to fourteen valves could be worked in series from 230-volt mains, and a more usual number of perhaps five together with the rectifier would absorb a potential drop of nearly 100 volts, leaving the remainder to be dissipated as heat in a resistance *R*. The adjustment of this resistance allows the receiver to run at any mains voltage from 110 or 250, and to assist this it is often external to the set. The latter is designed to run at the minimum voltage of 100 or 110, and for higher voltages resistance is added and may take the convenient form of a flexible cord built up from resistance wire. The difficulty of low high-tension voltage is overcome by the use of efficient pentode output valves able to deliver good volume even when the anode potential is less than 100 volts.

There is one type of mains rectifier other than the valve remaining in use. This is closely akin to obsolete chemical types and employs a pair of chemically prepared surfaces in contact, somewhat analogous to the crystal detector. The materials are copper and cuprous oxide suitably prepared, and hence the type is termed a copper-oxide rectifier, or sometimes a metal rectifier. Unlike its predecessors, however, the rectifier is reliable and highly efficient, and can be designed for wide ranges of current and voltage. A typical copper-oxide 'cell' will not operate safely above a certain potential of a few volts only, and the current density across the rectifying surface must not exceed definite limits or the temperature become too great. Higher voltages necessitate the use of several cells in series. These are built up of alternate disks of copper and of the oxide, and many of the copper ones are of larger diameter to act as cooling flanges. By the use of large-diameter disks currents of several amperes can be rectified, and the device is more convenient than the valve for large currents at low voltage such as might be required for accumulator-charging. As high-tension

rectifiers the need for a number of cells in series makes the rectifier as costly as the valve, but it offers the advantage of requiring no heater current and of being very durable if not overloaded. It is therefore well suited to the smaller or more portable class of receiver, or to the production of grid bias, for which it is widely used. Whilst quite practicable at higher voltages and currents the copper-oxide rectifier becomes more expensive than the valve, and so is increasingly less employed in the field of high-powered equipment or transmission.

Fig. 80 shows how four rectifiers can be connected to give

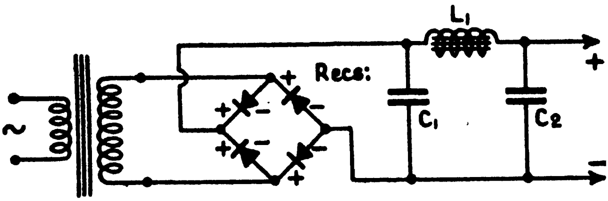


FIG. 80.

full-wave rectification from a single transformer winding or direct from the mains. This property makes them very suitable to universal use. An analysis of this circuit shows it to be equivalent to two full-wave rectifiers working in series. As a result the voltage built up across C_1 is approximately twice the alternating voltage input to the rectifiers, and the arrangement is termed a 'voltage-doubler' circuit. It provides an economical source of a voltage somewhat higher than that given by the mains or input transformer.

Recently small types of copper-oxide rectifier have been produced having low internal capacities and which can be used as detectors at the lower radio-frequencies. They will replace a crystal detector or the diodes used in multi-stage receivers. They are also found very convenient as rectifiers for use in series with direct-current measuring instruments to enable these to read alternating currents up to quite high frequencies, and special types are made for this purpose.

We have noted that many receiving or amplifying circuits will only work normally if fed from a high-tension supply of low internal resistance. An appreciable resistance at this point is common to the anode circuits of every valve used, and if for example the first and last valve of an amplifier contain such a

common resistance, potentials set up across it by the latter may be transferred back to the first grid. This is a form of reaction which may easily set the whole circuit into unwanted oscillation. Even in less serious cases the working of a circuit may be seriously impaired if energy passes from one stage to others through the coupling effect of the high-tension supply.

To overcome this a method known as 'decoupling' is employed in the design of most modern equipment. By its use the anode circuit of each stage is completed by a large condenser to its own cathode, as shown at C_2 in Fig. 81, whilst it receives anode current through the decoupling resistance R_2. C_2 and R_2 together form a simple filter. Alternating signal components at any but the lowest frequencies pass through the decoupling condenser C_2, and are rejected from the eliminator circuits by the resistance R_2. Any internal resistance of the eliminator will now be unimportant provided that it is small relative to R_2. A decoupling filter of this kind is fitted as a matter of course to any stage of a modern circuit which may need it. To simplify the diagrams, however, it is not shown in those which accompany later chapters, and the reader should assume it to be present whenever necessary. We can therefore assume that in all subsequent circuits the valve stages receive anode current from sources which are independent of each other and can cause no stray reaction. The sources might in practice be a single eliminator, but this will be treated as of zero internal resistance, because decoupling is assumed where necessary.

The design of decoupling circuits is mentioned here because it is bound up with that of anode supply. The larger both C_2 and R_2, the more complete the decoupling effect, and we should make both very high were it not for practical considerations. The condenser must withstand the full anode voltage in safety, and at the moment of switching-on, when the valve cathodes may not be hot, it may receive a still larger voltage. The size and cost of this condenser will therefore be important, and it will not usually be of larger capacity than necessary to present a negligible reactance at the lowest frequency amplified. In most cases a capacity of a few microfarads is ample. Many American receivers employ 0·1 microfarad as a standard figure, whilst in Great Britain from 1 to 4 microfarads is more usual. Only in exceptional cases such as that of television amplifiers will the capacity exceed 10 microfarads. We shall see later that

grid-bias circuits may also be decoupled, and the size of condenser necessary depends very much upon whether this is done, also upon the general design of the equipment and the number of stages concerned.

A limit is set to R_2 by the potential drop which occurs in it when anode current flows. This is of course simply calculated, being given in voltage by the product of the resistance R_2 in ohms and the current in amperes, or more conveniently of resistance in thousands of ohms and current in milliamperes. For a given high-tension voltage and working anode voltage, R_2 is thus definitely fixed. Should it not be high enough in relation to the internal resistance of the eliminator, an increase is only possible by raising the voltage output of the latter. As a rule resistances are needed, as shown in Fig. 78, to reduce the high-tension voltage supplied to the earlier stages of a receiver. It has become the usual practice to-day to combine these with the decoupling resistances, a single resistance R_2 being chosen for each stage which will give the desired potential drop. There is thus no need for a separate resistance network for voltage adjustment and decoupling as shown in Fig. 78, the two problems being tackled as a single item of design.

The construction of filter condensers is outside the scope of this book. In general they will be wax- or oil-impregnated paper-dielectric condensers having metal-foil electrodes. They can be treated as pure capacities, since their losses and inductance will be negligible at mains frequency. The only factor to be considered is that they shall be suitable to work continuously at the high voltage used. A modified construction of condenser has come into use, however, which differs sufficiently in electrical properties to need special mention. This is the electrolytic condenser, having a fluid or semi-fluid dielectric, and closely related to the secondary battery. If, for example, a battery of secondary cells equal to the total high-tension voltage were 'floated' across an eliminator in place of C_2 in Figs. 77 or 78, it would have the effect of maintaining the voltage steadily at its nominal value. Any drop in output from the eliminator would be made up by the battery, whilst any increase would be absorbed in charging it. The battery would possess a low internal resistance to alternating pulsations, which would be largely smoothed out.

The electrolytic condenser has a somewhat similar action.

It consists essentially of two aluminium electrodes, one of which usually takes the form of an outer can or container holding the electrolyte, and the second of a roll of aluminium sheet supported within it. The electrolyte is essentially a solution of ammonium borate, with boracic acid, borax, and other salts. These are dissolved in distilled water for the 'wet' type of condenser, or formed into a paste with glycerine for the 'dry' type. The condenser is polarized and must be connected to a circuit in the correct polarity, the outer can being the cathode and the inner foil the anode. After manufacture the condensers must be 'formed' by the passage of a large current, somewhat as an accumulator is charged. At first the current is considerable, but steadily falls until it reaches a low steady value of perhaps a quarter of a milliampere, at which it remains. This leakage current is necessary for the correct working of the condenser, and persists throughout its life. The condenser cannot therefore hold a charge for a long period, and has a relatively poor power factor. It is not suited to use in resonant circuits, for example, and is largely confined to eliminators and decoupling. The forming process is believed to coat the anode with a film of aluminium oxide which is intensely thin and forms the true dielectric, the electrolyte being the effective 'negative electrode'. The film becomes thicker with higher forming voltage, and it is found that the final effective capacity is inversely proportional to the forming voltage. A condenser of typical construction might, for example, have a capacity of 500 microfarads if formed and used at only 8 volts, as in grid-bias decoupling circuits. The same condenser, however, could be changed to one of 8 microfarads at about 500 forming voltage, when it would be suitable for use in an eliminator delivering 500 volts.

The electrolytic condenser is invaluable in the position C_2 of an eliminator, for several reasons. Firstly it is cheap and small, providing a large effective capacity in very handy form. By nature of its electrolytic action and polarity the condenser is also a voltage-limiter and partial rectifier, which are useful properties in an eliminator circuit. The highest voltage for which it can be designed is about 550, and any increase in voltage across it is dissipated as an increased leakage current. If prolonged under excess voltage, the condenser would heat up and might be damaged, but it is able to regulate temporary voltage surges such as those which occur at the moment of

switching on a receiver. By doing so it eases the strain on other condensers and components, which can in consequence be cheapened.

The condensers must not be connected in reverse polarity for long periods, as the resulting large current may destroy the oxide film. Their rectifying action allows them safely to by-pass a part of the alternating ripple, and smoothing may be more effective than with normal condensers of equal capacity, which would also be more costly. They must not be used on alternating current, and are unsuitable for the position of C_1 in Figs. 77 or 78. By forming a film on both electrodes it is possible to construct a 'non-polarized' electrolytic condenser, which whilst still unsuited to prolonged use on alternating current may be connected into a direct-current circuit either way round. In ordinary use at slightly below the maximum voltage the small leakage current constantly repairs the oxide film. Electrolytic condensers, particularly the wet variety, can thus withstand momentary overload, and are self-repairing after an occasional breakdown of the dielectric. These properties of durability, effectiveness, and low cost have made them an important factor in the improvement of modern mains-driven radio-equipment.

EXAMINATION QUESTIONS

1. Describe the action of a crystal detector. How does the response of a crystal detector vary with signals of different amplitude? Why are valves more generally used at the present time in preference to crystal detectors?

City and Guilds of London Institute. Preliminary Exam. 1937.

2. Why is it desirable to use a permanent magnet in the ordinary telephone receiver? What happens if the permanent magnet of a telephone receiver is replaced by a soft-iron core?

C. and G. of L. I. Preliminary Exam. 1936.

3. Describe three methods in which a thermionic valve can be used for detection. Explain the action of the valve in each case.

C. and G. of L. I. Preliminary Exam. 1934.

4. Why is a detector stage necessary in a wireless receiver? Explain clearly the purpose of this stage, and how this purpose is achieved. *Institute of Wireless Technology. June 1934.*

5. Compare the three best-known methods of detection from the point of view of faithful reproduction of telephony.

6. Explain the use of diodes and three-electrode valves in rectification. Give reasons for the employment of a diode for rectification in modern broadcast receiving sets. *Grad. I. E. E. 1935.*

7. Why is a filter composed of resistance and capacity necessary after any form of detector? What is its function, and what effects can it have on the quality of reproduction of telephony? Illustrate the use of such a filter in a simple crystal receiver.

8. Give a circuit diagram of a full-wave thermionic rectifier complete with smoothing circuit suitable for supplying high-tension direct current to a valve transmitter. *C. and G. Grade 1 1933.*

9. What are (*a*) a metal rectifier, (*b*) an electrolytic condenser, (*c*) a voltage-doubler circuit? Draw a sketch of a mains high-tension eliminator in which all three of these are used.

10. What is 'decoupling', and how is it applied in modern receiver design?

11. What is meant by the 'regulation' of a high-tension supply circuit? What effect upon the working of a receiver may be expected if the high-tension supply has high internal resistance, and what steps would you take to remedy the condition?

CHAPTER VIII

AMPLIFICATION OF ELECTRICAL OSCILLATIONS

We have studied the use of the valve as a detector of radio-signals, in which capacity it has gradually replaced the crystal and a variety of earlier devices on account of its superior stability and reliability; but a still more important use to which the triode valve has been put is that of amplification. It is this property that has made the design of efficient radio-receivers possible, whilst its applications in other scientific and industrial fields are enormous.

There are two chief methods by which valve amplification can improve the receiver. The earliest of these to be used is termed 'low-frequency amplification', or, in American circles, 'audio-amplification'. Here the valve is used to magnify the audible signals produced by a detector, so that these become louder in the telephones or can operate a loud speaker. Signals already received are made easier to hear or strong enough to work recording devices, and fresh signals previously too weak to be heard may become audible; but low-frequency amplification will not help the detector to rectify signals already too weak to operate it and so will not extend the range of a receiver greatly. Speech and music are made up of component frequencies ranging from practically zero to about 15,000 or more cycles per second, and for substantially perfect reproduction an amplifier should treat the whole of this range equally and without distortion. Practically, however, a somewhat smaller range is usually sufficient, and the reproduction of an average receiver may extend from 25 to 5,000 or 8,000 cycles per second, whilst retaining traces of higher and lower frequencies in decreasing proportion.

To increase the sensitivity of a receiver it is necessary to amplify the incoming signals before they reach the detector. This is the second method mentioned, and is referred to as 'high-frequency amplification'. Recently the American term 'signal-frequency amplification' has come to be used in cases where there might be confusion with stages working at some other high frequency, such as the intermediate frequency of the superheterodyne receiver. There are several advantages of ampli-

fying the incoming signals before detection. One of these is that signals too weak to operate the detector directly from the aerial can be brought up to a value at which they will be effectively rectified. Another is that additional tuned circuits can be introduced into the amplifier, thus improving the separation of stations on adjacent wavelengths; whilst yet another is the improvement in the ratio of signal strength to noises produced within the receiver itself. In high-frequency amplification the frequency to be amplified may lie between the limits of a few kilocycles up to hundreds of megacycles, and the technique will differ accordingly, but it is not usually necessary to amplify any wide band of frequencies at one time. The amplifier circuits can therefore be tuned to the particular frequency required at the moment, with a material improvement in efficiency.

In addition to these two types of amplification, which are easily distinguished in the average receiver, there are cases in which a very wide band of relatively high frequencies must be handled without distortion, as in television reception. There are also special cases where direct as well as alternating potentials must be amplified, and cases where distortion is permissible or even necessary, such as the frequency-doubling stages of transmitters. We shall deal with the subject of amplification from its fundamental aspect, indicating the practical circuit arrangements most suitable to various frequency ranges. These differ only in the form of circuit used to couple each amplifying valve to the next.

We have seen that if steps are taken to maintain the anode current of a triode valve constant, then any change in grid potential which takes place must result in a corresponding change in anode potential. If we assume that the change in grid potential shifts the operating-point along a straight portion of the characteristic curve, and also that absolutely no change in anode current is allowed to occur, then the change in anode potential would be strictly proportional to that of the grid. Any oscillatory potential applied to the grid would be reproduced in amplified form between anode and cathode, irrespective of wave form or frequency, and the condition would be one of perfectly distortionless amplification. This brings out the first essential point concerning amplification, which is that for no distortion of wave form to occur in a single amplifying valve, operation must take place along a part of the characteristic

curve which is perfectly straight. This applies of course to the grid-potential anode-potential curves, but since these are of the same form as grid-potential anode-current curves such as those of Fig. 38 it applies also to these. Only the central portion of these curves is approximately a straight line, and to confine operation to this we must not exceed a certain maximum oscillatory grid potential, and must employ a fixed grid bias which will place the point about which the grid potential fluctuates

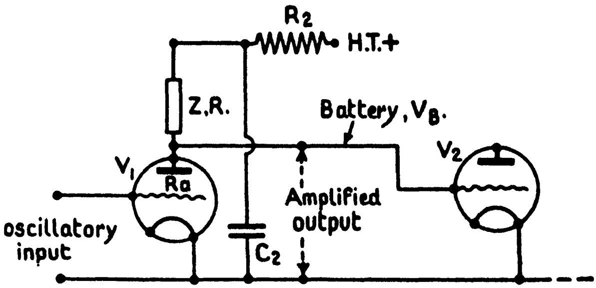


FIG. 81.

at the centre of this straight portion. It is clearly an advantage to select valves having the longest and straightest characteristics, and to employ high values of anode potential, as these lengthen the straight portion of the curve. Should the operating-point be allowed to encroach upon the curved portions near anode-current saturation or cut-off, there will no longer be strict proportionality between input and output potential changes, and the amplification will be accompanied by distortion.

Valve amplification depends upon the preceding simple effects, and its practical realization and defects arise from the methods used to maintain constant anode current and to transfer the amplified anode potentials to the grid of a following valve. Considering the first of these factors, it is impossible in practice to maintain absolutely constant anode current, and a near approach to this must suffice. It is obtained by the method mentioned in Chapter VII and shown in Fig. 81. An impedance Z is introduced into the anode circuit, which tends to oppose any change in the current through it, and maintains this approximately constant during rapid alternations in grid potential. This impedance may consist of resistance, inductance, or a mixed impedance containing both these together with capacity,

and its behaviour will differ somewhat in these various cases. The large condenser C provides a low reactance path from +H.T. to cathode, so that potential changes across Z can be regarded as also built up between grid and filament of V_2.

It is a necessary condition in the simple theory of amplification which follows that the H.T. source shall be of negligible impedance, and the condenser C helps to ensure this. It can be omitted when the resistance of the H.T. source is very small.

We will study first the action of a pure resistance R. If this be large compared to the internal resistance R_a of the anode circuit, then any change in the latter cannot produce a large change in the total current through R, and the necessary constant-current condition is approximately attained. The effect of changing grid potential on the valve can be regarded as producing a change in R_a, and when this occurs the potential across R_a due to the constant anode current must change also. Since the total potential across R and R_a in series is constant and equal to the high-tension voltage, any drop across R_a must be accompanied by an equal rise across R, which can be passed on to the following valve.

From the point of view of maximum amplification it would be desirable for R to be infinite. Were this the case we should obtain the maximum possible amplification, equal to the voltage-amplification factor μ of the valve, and the whole of the amplified potential would appear across R. The changes in current produced by variations in R_a would become infinitesimal as R became infinite, and so the constant-current condition would be realized. In practice there are other factors which make extremely high values of R undesirable, but the greatest amplification occurs when it is as great as possible consistent with these factors. This is borne out by the expression for amplification, which can be shown to be:

$$\text{Amplification} = \left(\frac{R}{R+R_a}\right)\mu.$$

Neglecting frequency considerations, the factor which limits R is the necessity for a sufficient mean anode current and potential to place the working-point suitably upon the valve characteristic curve; for we have seen that if an alternating grid potential of several volts is to be amplified without serious distortion a sufficient anode potential must be available so that

the characteristic curve is flat over a sufficient length. Also it happens that for most valve types the working portion of the curve becomes more nearly linear as the anode potential is increased, and so for minimum distortion this should be high. At this point it can be noted that exceptionally high values of anode resistance can be used when very small signals are to be amplified and distortion is not a very important consideration. This was frequently done some years ago and is convenient when receiving Morse signals only, when it enables increased amplification to be realized from any given valve. We will now consider the more usual case, however, when signals of greater strength are to be amplified without distortion, and estimate a suitable value of anode resistance to be used.

It is hardly possible to give a general expression for the best value of anode resistance since this is a compromise depending upon the performance desired. It is determined by one of several limiting factors, such as the maximum available high-tension voltage, the permissible degree of distortion, or the amplitude of the alternating signal potential which it is desired to build up across the anode resistance. In the case of triode valves the anode resistance will lie between the limits of the R_a and about ten times R_a, a usual value being from two to three times R_a. Higher values provide little increase in amplification but require very large high-tension voltages, whilst lower values can be shown to result in increased distortion and the presence of spurious harmonics in the output. For this reason the anode resistance should not be much less than $2R_a$, and it is bad practice to limit the amplification when necessary by reducing this resistance below R_a. Harmonic distortion may be defined as the total percentage of harmonics present in the output from an amplifier when a pure sine-wave potential free from harmonics is applied to the grid, or sometimes as the percentage of a particular harmonic so introduced. For a correctly designed triode stage it should not exceed 2 per cent.

In selecting an anode resistance the most usual procedure is to select an operating-point upon the characteristic curves of the type of valve to be used, such that the length of the straight portion on either side of this is sufficient to accommodate the maximum signal amplitude expected. The mean values of anode voltage V_a, of grid bias V_b, and of anode current I_a at which the valve will work in the absence of a signal can now be

read off from the curve. These must of course be within the safe rating of the valve, and must not exceed the values available. The negative grid bias should next be set to the selected value by means of a biasing battery or the other methods of biasing to be mentioned later. This bias must be negative, and must exceed the peak potential of the signal applied to the grid, for if at any time the grid is allowed to become positive it behaves as an anode and attracts electrons to itself to form grid current. This current represents a potential change between grid and cathode not proportional to the incoming signal, and will therefore modify the variations of grid potential, giving rise to distortion. There are certain amplifying systems in which grid current is allowed for and permissible, but in the simple case we are now considering it must never occur.

Having fixed the operating-point and grid bias, we have a certain known value of anode voltage and current at which the valve must work. Usually there will be a known value of high-tension voltage, also determined by the available source of supply, which we will call V. Then the actual anode voltage will be that of the supply less the potential drop due to the anode current flowing through the anode resistance. This is expressed in the equation:

$$V = V_a + I_a R \quad \text{or} \quad R = \frac{V - V_a}{I_a}.$$

Knowing V and V_a in volts and I_a in amperes, R in ohms can be found. Any value of R worked out from this expression will place the valve at a suitable point upon its curve, but the value chosen must also be examined from the amplification standpoint. Should it come out to be less than about twice R_a, it will be necessary to select fresh conditions which allow of a higher value. This can be done by increasing the high-tension voltage V so that for given values of V_a and I_a R will become greater, and a little consideration will show that V must always be considerably greater than V_a for high values of R. Should it be impossible to increase V, then there is no alternative but to select a new working-point upon the valve characteristic at which the anode current and potential are lower, and R consequently greater. This must involve a reduction in the maximum signal amplitude which can be amplified with a minimum of distortion, and hence we see that for the valve to amplify at

high levels of amplitude the high-tension voltages used must be high. In practice it is always desirable for V to be as high as possible, since it results in a straighter characteristic curve with less harmonic distortion, and in higher working values of R and consequently a better approach to the maximum possible amplification of which the valve is capable. The equation makes it clear that R and V must be increased together if the valve is to work at any given point on the curve, and it is for this reason that very high anode resistances necessitate an impracticably large high-tension voltage.

An advantage of a pure resistance used in this manner to maintain the anode current approximately constant is that theoretically the amplification should be independent of frequency. At low frequencies this is nearly the case, and hence the resistance-coupled amplifier is largely used in the range between zero and some 10,000 cycles per second, which comprises the bulk of speech and musical frequencies. At the highest of these, however, the resistance must be regarded as a mixed impedance comprising resistance and capacity in parallel, because the small capacities due to the anode-cathode path of the valve and the grid-to-cathode of the following valve or load circuit appear across the anode resistance. These and the capacities of wiring and components to earth combine to form a parallel path across the anode resistance, and although only a few micromicrofarads they provide a reactance decreasing with frequency and becoming comparable to the anode resistance. Thus, as we attempt to amplify frequencies above the lower audible range, the impedance of the anode circuit becomes increasingly lower than that of the original anode resistance. It can be shown that the effect of a mixed impedance as anode-circuit load is similar to that of a resistance having the same value in ohms at the frequency under consideration, and amplification occurs in the same manner and to the same degree. But the effect of these shunt capacities lowers the total impedance so greatly at high frequencies that resistance coupling ceases to be effective. The kind of curve given by a simple resistance in the anode circuit is shown at (*a*) in Fig. 82. It is not drawn strictly to scale.

In early equipment attempts were made to employ resistance coupling at radio-frequencies, and to use it in front of the detector of receivers working at the longer wavelengths. The

amplification obtainable in this manner is a very small fraction of that possible from the valves, but may not be entirely useless. When 'aperiodic' amplification (independent of frequency) is desired at frequencies above the audible range, such as in television or radio-frequency stages, three modifications of the simple resistance-coupling are in use. Firstly, it is an improvement to employ valves of low R_a which can be effectively

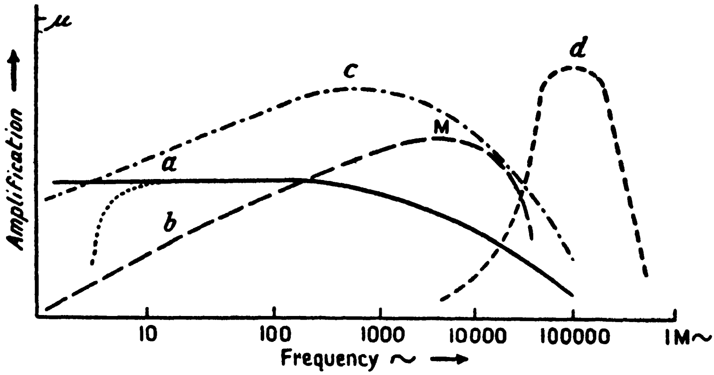


FIG. 82.

coupled by low values of anode resistance. The reactance of stray capacities will then be higher relative to the anode resistance and will have a smaller shunting effect upon it. In addition it is obviously desirable to employ valves and components having minimum self-capacities, and modern methods of design have assisted materially in that direction.

The next step is taken by the insertion of an inductance of comparatively small value in series with the anode resistance. Now, the reactance of an inductance increases proportionately with frequency, and since it is placed in series with the resistance the combined reactance will also increase with frequency. But the reactance of the shunt capacities falls proportionately with frequency, tending to lower the combined reactance. The effect of the inductance and the capacity are thus opposite in their effect upon the total impedance of the anode circuit, and by suitable choice of values can be made to compensate for each other over a very wide range of frequency. Reference to Chapter II will help to explain this behaviour. There it is shown that the impedance Z of such a circuit is given by the expression:

$$Z = \sqrt{\{R^2 + (L\omega - 1/C\omega)^2\}} \text{ ohms.}$$

And this will remain $Z = R$ simply, as long as $(L\omega - 1/C\omega) = 0$. Now this will only occur at one particular frequency. Let this frequency be selected near the maximum at which the amplification is to be maintained, then $Z = R$ both at very low frequencies, when both LC and C tend to zero, and also at the selected maximum frequency. The amplification will be the same at these two extremes. At intermediate frequencies the cancellation between inductive and capacitive reactance will not be exact, but since both quantities are small compared to the anode resistance itself and are varying in the opposite sense, their difference will never change the total impedance as much as would either separately. In this way a reasonably uniform amplification is possible up to frequencies of more than a million cycles, but only by the use of a relatively low total impedance, resulting in 'under-matching' of the valve alternating-current resistance and consequently poor amplification per stage.

When circumstances do not demand the amplification of the lower frequencies, or when we wish to deal with a particular range of high frequencies without the necessity of equal treatment for those approaching zero cycle per second, it becomes possible to omit the anode resistance and to rely entirely upon an inductive reactance. In practice the amplifier then becomes 'choke-coupled', and the anode-circuit impedance consists of a large inductance or choke coil having only its own resistance acting effectively in series with it. The frequency characteristic of choke coupling rises, the amplification increasing as frequency is raised, up to some maximum value at which the shunting effect of circuit and electrode capacity again becomes important, after which the parallel capacitive reactance gives rise to a falling characteristic. The curve of amplification against frequency is of the form shown at (*b*) in Fig. 82, rising to a maximum at M and followed by a fairly sharp 'cut-off'. If the inductance used is large, its reactance will exceed that of a practical anode resistance, and the peak amplification at M will approach quite nearly to the μ of the valve. An important advantage of choke coupling is that, since there is little resistance in the anode circuit, the anode potential is but little less than that of the high-tension source. which results in a very useful economy. It is no longer necessary to employ a high-tension voltage several times that actually required at the anode, in order to obtain good amplification.

The form of choke to be used will depend upon the frequency range to be amplified. One important case is the low- or audio-frequency amplifier, which must operate with little loss of efficiency down to a few cycles per second, and in this case an iron-cored inductance is used having a value of between 10 and 100 henries. This must be designed with sufficient magnetic core material to avoid magnetic saturation by the mean anode current flowing through it, and the core must be of laminated soft iron or material which will not show appreciable hysteresis losses up to the highest desired frequency. If the inductance of this choke is high it will possess a reactance of several thousand ohms at the lowest audible frequencies, and will therefore represent a reasonably efficient anode-circuit load. The response curve will resemble Fig. 82 (*c*), falling as before at very high frequencies owing to the self-capacity of the choke.

For the amplification of a band of higher frequencies which do not extend down into the lower audible range a choke of lower inductance is preferable. This need not contain an iron core, and in fact such would be undesirable as it would introduce losses which would lower the effective impedance of the choke. An air-cored coil will remain efficient up to a very high limit, and if the self-capacity be kept low and the inductance is also low, then the shunting effect of stray capacities will not limit the amplification seriously. The value of inductance is selected so that its reactance will be two or three times R_a at about the middle of the frequency range to be amplified, and a peak amplification of about 70 per cent. of μ will result. As the reactance of the choke rises with increasing frequencies, there will be a level or slightly increasing amplification until the losses due to capacities begin to be felt, whilst at lower frequencies the efficiency slowly falls to approach zero at a minimum frequency. Here the reactance of the choke is very low and has little coupling effect, but its residual resistance may still give rise to slight amplification.

The characteristic of a choke-coupled stage as outlined above is quite flat, provided that the resistance or damping of the anode circuit remains substantial. It is obvious, however, that the combination of inductance and the stray capacities will form a resonant circuit, and that this may tune to a point within the useful range of the amplifier. The inductance itself will have self-capacity and together with the valve electrode

and stray capacities may resonate at a frequency well below the maximum of the amplifier response, and, if the circuit resistance and damping are not too great, this resonance may exert a marked influence upon the amplification. In Chapter II we saw that at its resonant frequency a combination of inductance and capacity behaves as a pure resistance, which may be of quite high value, and which can therefore behave as an efficient anode-circuit impedance. Such resonance in the choke tends to impart a comparatively sharp maximum to the amplification curve.

Modification along these lines brings us to a further and very efficient form of amplifier for limited frequency ranges, in which the anode impedance Z is simply a tuned circuit of low resistance and damping. It is simply necessary to reduce the inductance value to such an extent that the inductive reactance is very much less than R_a and produces little amplification, and to tune this to resonance at the desired peak frequency either by the stray circuit capacities or by an added condenser. This may be a variable condenser, which will enable the optimum frequency to be adjusted over a wide range. The anode impedance Z now behaves as a pure resistance at resonance, as will be explained in Chapter X, and the circuit is that widely used and known as 'tuned-anode coupling'. Except at very high radio-frequencies it is possible for the equivalent or dynamic resistance of a good tuned circuit to exceed the R_a of average valves, and so efficient amplification is possible. As the ratio of capacity to inductance is increased, the dynamic resistance of the circuit clearly falls, and the tuning becomes sharper or more critical. The amplification characteristic curve then becomes very similar to that of the tuned circuit, having a sharp peak of maximum amplification and falling off sharply for adjacent frequencies, and is thus ideal for radio-frequency amplification of incoming signals, since it contributes to the total selectivity of the receiver. Such a curve is shown at (d) in Fig. 82, the resonant frequency being about 100,000 cycles.

We have now taken note of the possible forms of the impedance Z which must be introduced into the anode circuit of any valve in order that amplification can occur, noting that this can be a pure resistance, a choke of various types, or a resonant circuit, and that all three are modified at very high frequencies by the effects of stray circuit capacities. When alternating potentials are applied to the grid an amplified copy of these

occurs across the anode impedance, and must be passed on to a following valve or circuit. We must now decide how this can be done in practice, bearing in mind that we are dealing at present with pure 'voltage amplification', or the case where our object is to magnify potential changes. These are to be transferred to the grid of a following valve or to some other potential-operated device, for were they to operate telephones, which require appreciable power, the case becomes somewhat different. We shall discuss power amplification in a later chapter, and meanwhile assume that the grid-cathode circuit of the valve following our amplifier absorbs no power. This is not strictly true, but is exact enough for most purposes if the grid is negatively biased so that no grid current flows. In that case the grid can be regarded as an infinite resistance, although a more strict value would be given by μR_a, generally a very high figure. It must be remembered that the following grid-cathode circuit has capacity, however, which may be appreciable at high frequencies, and which acts as a portion of the total capacity in parallel with the preceding anode impedance.

In Fig. 81 an amplifying valve is shown in which a connexion is taken directly from the valve anode to the following grid. Since the filaments or cathodes of all the valves in an amplifier are taken to a common earthed conductor they are necessarily at identical potential as far as the signal variations are concerned, and any potential changes between anode and cathode of V_1 will be transferred as variations between grid and cathode of V_2. We need only consider therefore the coupling between anode and grid, bearing in mind that the other side of the circuit is common in all the cases under discussion. Now in the circuit shown potential variations are clearly transferred from V_1 to V_2, but since the grid of V_2 is directly joined to the anode of V_1 it will be at the full positive anode potential with respect to cathode, an impracticable operating condition. One method of overcoming this is to connect a battery between anode and grid. This must have a voltage V_b slightly greater than that on the anode of V_1 and is connected in the reverse potential sense, so that the resulting potential on the grid of V_2 is $V_a - V_b$. Since V_b slightly exceeds V_a there is a small negative potential at the grid, providing a suitable negative grid bias. This arrangement is termed a 'battery-coupled amplifier', and is a special case of an important class termed 'direct-coupled amplifiers'. Many

other methods have been used to obtain the correct grid potential, such as by isolating the respective cathodes by a large condenser having low reactance to the signal potentials. On applying a high positive potential, slightly in excess of V_a, to the cathode of V_2 a small resultant negative grid potential is provided.

However the biasing may be arranged, direct-coupled amplifiers have one unique property in common. Since there is a direct connexion, all variations at the anode of V_1 are transferred to the grid of V_2 without any frequency discrimination, and this applies equally to a frequency of zero, namely a direct current. For this reason the form of circuit is also termed a 'direct-current amplifier', since it can be used to magnify very slow changes of potential which are in no sense alternating. The anode impedance used for V_1 will naturally be a pure resistance which is invariable with frequency, and the circuit has then a uniform amplification with an absolute minimum of distortion from zero up to a very high frequency determined by the stray capacity effects. For radio-reception this is not usually needed, but it is very valuable for other special purposes such as television. The direct-coupled amplifier is less simple to design and operate than others, as the exceptionally wide response tends to promote instability, and elaborate precautions are necessary to ensure steady supply voltages. Slight changes in these are amplified by the circuits themselves and may result in instability at very low frequency or in gradual loss of the correct bias settings. It is very difficult indeed to employ more than two or three stages of direct-coupled amplification, and the practicable gain is limited. The circuit is for these reasons not usually employed except in those cases where response to the lowest possible frequencies or to direct potential changes is essential.

When direct coupling is not necessary these difficulties can be simply overcome if a condenser C_1 be inserted between anode and grid, as shown in Fig. 83. This serves to block out all direct potentials from reaching the grid whilst presenting a low reactance to the alternating signal potentials. The grid is thus isolated, and, since it is essential to provide it with bias and a path through which possible grid current can flow away to cathode, a grid leak is employed. This consists of a high resistance r_1 which joins the grid to a negative biasing battery or

negative filament as shown. The arrangement is termed 'resistance-capacity' coupling, and is one of the two types most widely used in radio-reception.

The action of the grid condenser is to prevent any direct potentials from reaching the grid. It must therefore be of the highest insulation, and mica dielectric or the best oil filling is necessary. Wax-impregnated-paper condensers are sometimes

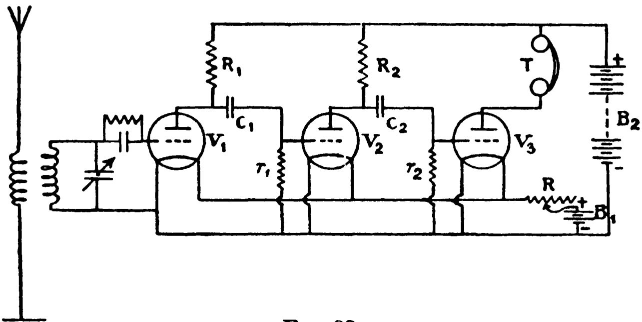


FIG. 83.

used but may not be entirely satisfactory. At the same time the condenser must pass on alternating potentials without loss. Were the grid-cathode circuit of infinite impedance no difficulty would exist, but we have seen that this is not strictly so, and also there is the grid leak in parallel with that circuit and having finite resistance. The reactance of the grid-coupling condenser now forms a potentiometer in conjunction with the grid leak, and the potential reaching the grid will be reduced by any drop which occurs in the condenser. At medium and high frequencies the reactance of the condenser may be negligible and the amplifier behave as if it were absent, but towards the lower extreme the reactance rises inversely with frequency until at a few cycles per second any practicable value of condenser offers such a high reactance that amplification becomes low. At zero frequency the reactance is infinite and no amplification can occur. The dotted addition to curve (*a*) in Fig. 82 shows the effect of the coupling condenser upon response. It is thus best to use the largest capacity possible and also to keep the leak resistance high if the low-frequency response is to be a maximum.

Unfortunately a compromise is necessary here because a large condenser cannot discharge quickly through a high-resistance

leak, the combination being said to possess a large time constant. If in operation the grid momentarily becomes positive through the arrival of a signal impulse which exceeds the grid bias, the coupling condenser will charge up and will maintain the grid positive until this charge can leak away through the grid leak. While the grid is positive distortion will occur and the amplifier may even cease to function, hence it is necessary that the positive charge leak away so quickly that the ear will not detect the occurrence. This implies a time constant of a small fraction of a second, and limits the grid condenser and leak values which may be employed. Practical values for low-frequency work are from 0·1 to 0·01 microfarad and 100,000 to 500,000 ohms, whilst if the condenser be reduced still further the leak may occasionally be raised to a few megohms. These values impose a limit to the amplification of the lowest frequencies, and in general the resistance-capacity-coupled amplifier shows a falling curve below a few hundred cycles, and a virtual cut-off below some figure between 20 and 100 cycles per second. When amplifying higher frequencies than the audible range this limitation does not of course occur. It is also possible to employ large condensers and leaks giving better low-frequency response if steps are taken to avoid any possibility of the grids becoming momentarily positive. This implies weak signals and high values of grid bias.

Choke-coupled circuits of either type also employ the condenser and leak to transfer amplified potentials to the following valve, and the considerations governing condenser and leak values are similar. Space will not permit of the complete calculation of the values of these components here, but in general when designing a coupling the anode impedance will be decided first. The valve and operating potentials having been selected to conform to the amplitude, degree of magnification, and frequency range desired, no great difficulty will be found in selecting a resistance, choke, or resonant circuit which presents an impedance of from 2 to 10 times R_a over that range. This can be simply calculated or measured. A value of condenser and leak suitable to the frequency range can next be selected, and examined to see if the losses it introduces are excessive.

The impedance of the condenser and leak in series will be given by the expression $\sqrt{\{R^2+(1/C\omega)^2\}}$, and hence the propor-

tion of the signal potential between anode and grid which reaches the grid is given by

$$\frac{R}{\sqrt{\{R^2+(1/C\omega)^2\}}},$$

which enables the loss due to the coupling condenser to be found for any particular frequency. Having selected values for these components which do not cause excessive loss within the desired range and have not an excessive time constant, it must not be forgotten that the condenser and leak in series are in effect in parallel with the anode impedance Z, and will modify its value. Their effect must be allowed for by an increased value of Z if necessary. This fact will be apparent from the diagram, if it is remembered that the positive high-tension line is at cathode potential for alternating E.M.F.s because the two are joined together by the low resistance of the high-tension battery or eliminator circuit, or by a very large condenser C.

Attention must now be given to the second method by which changes in anode potential can be transferred to the grid of a following valve. This is by inductive coupling, when the amplifier is termed 'transformer-coupled'. The method is only applicable to cases in which the anode circuit contains an inductive impedance, such as a choke or tuned circuit. This is then treated as the primary of a transformer, and is provided with a secondary winding inductively coupled to it, the two ends of this secondary winding being taken to the grid and cathode of the following stage. Clearly the essential conditions of coupling are fulfilled in this case also. The secondary winding is completely isolated from the anode circuit, and so the grid receives no unwanted biasing potentials, whilst a closed circuit from grid to cathode is available through which the grid may be correctly biased.

Fig. 84 (a) shows transformer coupling applied to a high-frequency amplifier. Here the high-frequency transformer consists of two air-cored inductances placed a suitable distance apart, such as two solenoids wound on a common former. The spacing between them of course determines the tightness of coupling. Three possibilities now exist. In the first case the transformer is untuned and the two windings closely coupled, the frequency characteristic being similar to that of choke-capacity coupling. Secondly, the primary or secondary (usually

the latter) is tuned and the coupling between the windings reduced, the performance then being very similar to a tuned anode circuit. In the third case both windings are sharply tuned by variable condensers and the coupling reduced to an extent which will result in a single sharp resonant peak, as described in Chapter IV and shown in Fig. 29 (*b*). It is also possible to couple the windings by a common impedance, as

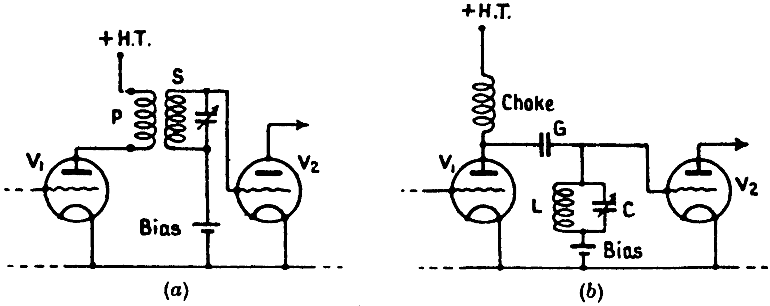


FIG. 84.

shown in Fig. 30 of that chapter. These arrangements result in increased selectivity, and are therefore useful in modern selective receivers and in the intermediate-frequency amplifier of the superheterodyne.

In the case where one winding only is tuned, or both are non-resonant, there is no objection to the use of a larger number of turns in the secondary than in the primary winding. The transformer then acquires a step-up ratio, and the voltage across the secondary would be increased in proportion to the turns ratio if there were no load across it. This should clearly lead to improved amplification. and does so in practice when the valves used have low R_a and high grid impedance. By analogy with resistance coupling, the amplification of a tuned-anode-coupled stage of dynamic (or equivalent resonant) resistance R' will be expressed by:

$$\text{Tuned-anode amplification} = \mu\left(\frac{R'}{R_a+R'}\right) \text{ or } \frac{\mu R'}{R_a+R'}.$$

On substituting a transformer coupling of turns ratio T, and having a tuned secondary winding L_2C_2, of dynamic resistance R_2', it can be shown that this represents a load of R_2'/T^2 to the valve anode circuit. R' must now be replaced by this expression, when, bearing in mind that amplification is also increased

by the step-up ratio T, the following expression gives the overall amplification of the stage:

$$\text{Transformer-coupled amplification} = \mu \times T \times \left(\frac{R_2'/T^2}{R_a + R_2'/T^2}\right).$$

For maximum power output from a valve, the internal resistance should equal the effective external resistance, or $R_a = R_2'/T^2$, when $T = \sqrt{(R_2'/R_a)}$ or the root of dynamic resistance over valve internal resistance. This formula gives the most suitable transformer ratio for a valve of internal resistance R_a. Notice that a condition for maximum power output has been assumed, since this minimizes any loss of effective amplification due to the loading of $L_2 C_2$ by the grid of the following stage, seldom an entirely negligible quantity.

The maximum possible amplification is found by substituting $R_a = R_2'/T^2$ in the above expression, when

$$\text{Maximum amplification} = \tfrac{1}{2}\mu \times T = \frac{\mu}{2}\sqrt{\left(\frac{R_2'}{R_a}\right)}$$

and it becomes clear that the ratio $\mu/\sqrt{(R_a)}$, which is related to mutual conductance, is a measure of the high-frequency amplifying properties of any valve. It measures the amplification obtainable from that valve under the best possible coupling conditions. Using low-impedance types of valves, amplification would be improved by the use of transformer coupling, but it so happens that many modern types of screen-grid and pentode valves have such characteristics that the optimum transformer ratio is nearly unity, and there is little to be gained by the use of an untuned transformer. Practically, therefore, the choice between tuned-anode and transformer coupling is often one of convenience. The former has the advantage of simplicity, which may be important when a receiver has to work over a wide range of wavelengths and employs switched coils, whilst the latter will in most cases be rather more efficient.

Mention might here be made of a form of coupling termed 'tuned-grid', which is in effect a modified form of tuned-anode coupling having certain of the advantages of the transformer. This is shown in Fig. 84 (*b*) and will be seen to employ a choke or resistance in the anode circuit of V_1 coupled by a condenser to a resonant circuit which replaces the grid leak. Since this circuit and the anode choke are effectively in parallel and the impedance of the latter should be considerably the greater, the

impedance matching approximates to that of the tuned-anode arrangement, but it is found more convenient because the grid is connected to cathode by a circuit of low resistance. This makes for stability, and it is also a convenience to have one side of the tuned circuit at cathode (earth) potential.

The case of low-frequency transformer coupling within the audible range is more advantageous, since here the transformer can employ a laminated iron core and coupling is

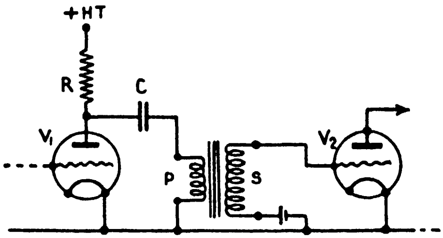


FIG. 85.

extremely tight. It is now possible to use a large step-up turns ratio so that the secondary voltages exceed the primary several-fold and amplification is very materially increased. It is difficult to design a transformer which will transform energy from primary to secondary with complete uniformity over such a range as from 20 to 10,000 cycles per second when the turns ratio is very high, and hence for a minimum distortion this ratio seldom exceeds 3:1. If, however, a degree of falling-off at the high and low ends of the scale can be tolerated it becomes possible to employ higher ratios, of from 5:1 up to perhaps 10:1. In each case the total amplification from the valve and coupling is increased by a factor substantially equal to the transformer ratio. There is thus a considerable advantage in the use of low-frequency transformers, particularly when large amplification is desired from the minimum number of valves and a slight degree of distortion can be allowed.

In Fig. 85 is illustrated a transformer-coupled stage, and notice should be taken of the 'shunt-fed' primary connexion. In early equipment it was usual to insert the transformer primary into the actual anode circuit, as shown for V_1 of Fig. 86, when the whole anode current flowed through it. This demanded a bulky transformer having a large enough core to prevent magnetic saturation. Now, for uniform response to the

lower frequencies the transformer primary must maintain a reactance well in excess of R_a, and this implies a high inductance, of perhaps 100 henries. This inductance can be obtained from a primary winding of fewer turns if the permeability of the core can be increased, which is made possible by the use of modern types of high-permeability alloys. Quite a small current will be sufficient to saturate this type of core, however, and in any

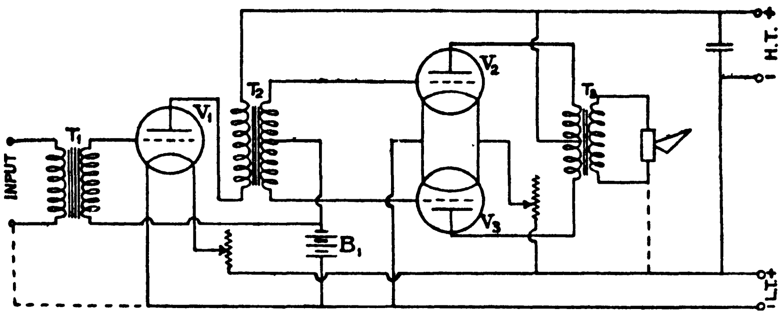


FIG. 86.

case the inductance is reduced by the presence of a permanent magnetic flux through the core. The circuit shown overcomes this difficulty, since the direct anode current does not flow through the transformer primary but through the feed resistance R. Signal potentials built up across R will be transferred through the blocking condenser C to the transformer primary, which is effectively in parallel with R for alternating potentials, and then inductively transferred at increased voltage to the secondary winding. As in resistance-capacity coupling the capacity of C must be sufficient to present a negligible reactance at the lowest wanted frequency, but as the reactance into which C feeds is high this presents less serious difficulty. Resonance is possible at a low frequency between C and the transformer primary inductance, and this effect is sometimes made use of to improve the amplification near the lower limit of frequency.

We have now outlined the chief methods by which amplifying valves can be coupled together and indicated their salient properties. From time to time we shall note instances in which the different methods are employed in practice. So far reference has only been made to the use of triode valves, which were for

many years the only ones in use, but the basic coupling circuits remain the same when multi-grid valves are employed. Only the component values and the working potentials need to be changed in order to suit the characteristics of the new valves, chief of which is usually a high value of R_a.

Whilst studying valve types we saw that the screen-grid has largely replaced the triode in high-frequency circuits because of its reduced tendency to self-oscillation and the greater amplification possible. The coupling of screen-grid valves is similar to that of triodes except that a very much higher value of anode impedance is necessary to match the very high R_a of the valve. At the shorter wavelengths it is not possible to attain the optimum value of impedance from any ordinary tuned circuit, which will lie between 50,000 and 250,000 ohms equivalent resistance at resonance and cannot exceed the R_a of valves ranging from about 250,000 ohms up to over a megohm. Thus the amplification increases as the circuit impedance is increased, and the latter is maintained as high as possible by low-loss design and a high ratio of L to C. The impedance of valve grid circuits at high frequencies is also of the same order as R_a, and the best ratio for transformer coupling will not be far from unity. Consequently transformer coupling is not widely used except at the longer wavelengths or for secondary reasons of circuit design, such as improved selectivity.

Good amplification and coupling efficiency can be attained by the tuned-anode or tuned-grid arrangements, which are widely used on account of simplicity. Selectivity can be improved if the valve anode be tapped down to an intermediate point upon the tuned circuit inductance, thereby reducing the load or damping of the circuit by the valve. Similarly, it is sometimes desirable to tap down the grid of the following stage, thus reducing the damping caused by grid current or capacity. This is particularly necessary when the following valve is a detector, as (with the exception of the anode-bend variety) detectors always draw considerable grid current and damp a circuit heavily. Such tapping-down to improve the impedance match does not necessarily reduce amplification, as might be expected at first sight, since a larger oscillatory potential will be built up across a lightly damped circuit, and this effect may more than compensate for the smaller proportion of this tapped off to the grid.

Recently it has been found preferable to employ pentode-

type valves for radio-frequency amplification. Radio-frequency pentodes are similar to screen-grid valves in general construction but have the added suppressor-grid electrode. It has been found possible to attain a better ratio of μ to R_a from these valves, together with other minor advantages such as a straighter curve and larger safe signal input without distortion, and as a result of these factors performance is still further improved in all respects.

In the field of low-frequency amplification the pentode has also found favour. Its main application as a power amplifier will be mentioned later, but as a voltage amplifier it is found to yield greater amplification per stage and can respond to higher frequencies than the original triode. For limited input amplitudes the screen-grid is also used as an efficient amplifier for audible frequencies. Resistance or impedance-capacity coupling must be used to couple these high-impedance valves, since transformers of sufficient inductance to match their very high R_a are not practicable. High anode voltages thus become necessary, and the main field for such technique is in high-gain amplifiers for television and similar uses, when it is found possible to amplify uniformly a wider band of frequencies at reasonable efficiency than can be done with triodes. In general work the triode coupled by a good transformer is still widely used, since the frequency range is ample for most purposes and the amplification per stage as high as can be easily obtained in any other manner. There is a tendency to-day, however, for low-frequency amplification to be omitted from receiver designs, since the very high efficiency of the pre-detector stages enables rectification to take place at high amplitude levels, when the detector can be directly followed by a power amplifier stage. This method of design is commercially convenient, as it places the amplification where it is most needed for reasons of sensitivity and selectivity, whilst keeping it down in the low-frequency section, which is the principal seat of hum and distortion.

When discussing oscillators we saw that advantages exist in the use of a pair of valves in push-pull. Push-pull operation is equally possible in radio-frequency or low-frequency amplification, and is resorted to when the cost of the additional valves and components is justified by the improved performance. Push-pull radio-frequency stages are mainly used in transmission or at very high frequencies. In the former case the advantages

are similar to those described in the case of oscillators, such as the elimination of stray radio-frequency potentials from the supply circuits, whilst in receiver design the reduction in the capacities due to the valve electrodes across the tuned circuits may help materially towards efficiency.

In the low-frequency field push-pull has very important special properties, the chief of which concerns increased power output and will be explained in the later chapter on faithful reproduction. Fig. 86 shows the circuit of a push-pull transformer-coupled power amplifier. As a voltage amplifier the push-pull stage reduces distortion, because even harmonics produced as a result of curvature of the valve characteristics are cancelled in the differentially wound transformer primaries. Also the two valves can clearly accept double the signal amplitude without overloading, or will each work over a reduced portion of their curves and with consequently less distortion when a given input is shared between them. An additional convenience is that alternating-current hum remaining in the high-tension supply largely cancels out in the anode circuits of a push-pull stage, whilst since the mean anode current of the two valves remains constant under all conditions there is less possibility of fluctuations in the high-tension supply adversely affecting other stages. The combined result of these properties of push-pull operation is reduced distortion and a more easily stabilized amplifier, less affected by supply variations.

Mention should be made of a modified arrangement similar to push-pull and in which two valves are used to share the signal input potentials between their grids in opposite phase, the anodes being joined together. Such circuits are sometimes termed 'push-push', or in a particular form 'para-phase', amplifiers. In the latter two complete amplifiers may be used to provide parallel paths for the same signal but in phase opposition. Since both work from the same supplies, any variations in load are self-compensating. They share the advantages of push-pull in stability and freedom from certain forms of distortion, but have not the fundamental property of balancing out even-harmonic distortion or alternating-current hum from the high-tension supply. An advantage is that they can be resistance-capacity- or direct-coupled when necessary to provide a specially wide frequency response, whilst the reversal of phase necessary for push-pull working demands the use of centre-

tapped transformers in the manner shown in Fig. 86. Resistance-coupled push-pull can be achieved by certain unconventional arrangements, in one of which a phase-reversing valve with an amplification of unity is used to feed the grid of one of a pair of valves in opposite phase to the other. In another the amplified signal in the anode circuit of one valve forming a push-pull pair is used to provide a reversed-phase potential for application to the grid of the other. These circuits, whilst offering some of the advantages of true push-pull, are in reality different in operating principles.

So far it has been assumed that an amplifying valve is operated about a point near the centre of the straightest portion of its characteristic curve, and this has been stated as the necessary condition for a single-valve stage to amplify with minimum distortion. This linear operating condition has come to be known by the American designation of 'class-A' amplification. It has become clear, however, that this is not the only condition in which valves can amplify effectively, and that others may be preferable when distortion is tolerable or necessary, or when valves are used in push-pull. We shall see presently that in push-pull power output stages greater power can be obtained if the valves are biased to a very low value of anode current by the use of high negative grid bias. A greater signal amplitude can then be applied to the grids with a consequently larger output, and the harmonics introduced into the output by allowing the valves to work over curved parts of their characteristics are largely cancelled out by the push-pull connexion. In this method of operation no grid current is allowed to flow, and the system has been termed 'quiescent push-pull' in Great Britain. It could be clearly applied to voltage amplification also, but since little is to be gained by so doing, and the quality of reproduction suffers, linear push-pull is to be preferred in that case.

A valve biased to substantially zero anode current is said to be working under 'class-B' conditions. This too was originally an American term, and in Great Britain a distinction is made between the quiescent condition, explained in the last paragraph, and the condition here termed 'class B', in which sufficient signal amplitude is applied to cause grid current. It is possible to drive the grid of a class-B stage positive to such a degree that the anode current rises to saturation on the positive peaks, and when this is done the whole length of the characteristic is being

used from top to bottom bend, and the power or voltage output becomes very great. When grid current is flowing the grid impedance becomes relatively low, and power must be supplied to maintain the positive potential peaks at their true proportion. This power must come from the preceding stage, and so we see that class-B operation is essentially a power problem rather than one of voltage amplification. The chief advantage of the system lies in increased power amplification, whilst there may be no increase at all in the potential variations obtainable from the anode circuit over those which must be applied to the grid. In push-pull form the circuit has valuable properties for the driving of loud speakers and power-operated output devices, and we shall consider it further when dealing with the output stage of receivers.

Class-B voltage amplification results in considerable distortion and is of little value when the wave form of a signal must be unchanged. The value of a single valve working in this manner lies in the fact that the distorted output can be analysed as the sum of the original input frequencies and of higher harmonics of these. The second and third harmonics may represent quite a large fraction of the total output, and we have seen when considering continuous-wave transmission that they can be filtered out and usefully employed. This is the basis of frequency-multiplication by valves, where the class-B condition favours a maximum output of the desired harmonics.

When valves are used for this purpose it is possible to operate them under a still more efficient condition termed 'class C', in which the negative bias is still further increased up to about double the value necessary to cut off the anode current. Still greater amplitude must now be applied to the grid, and no anode current will flow until the positive half-cycles of input potential exceed about half the grid bias. After this, and if the input be sufficient, the anode current reaches saturation momentarily during the positive input peaks, and the output will take the form of very brief but intense impulses at the input frequency. This wave form is still more distorted and contains even larger harmonic energy than in the class-B condition. Also, owing to the long cooling periods which the valve experiences between each pulse of anode current, the output power can be still further increased without risk of damage to the valve, and the largest possible output obtained.

If a resonant circuit tuned to the input frequency or a harmonic thereof be placed in the anode circuit of a class-C stage, it will receive impulses at regular intervals which will throw the circuit into oscillation at its natural frequency. This kind of excitation is akin to shock excitation of a circuit by atmospherics and is sometimes termed 'flick' impulsing. The important point is that the wave form of the oscillations built up will be very nearly sinusoidal, whatever that of the impulses may be, owing to the flywheel effect of the resonant circuit, and so a useful output at a harmonic frequency is obtained. It is for this reason that class-C frequency-multipliers give an output similar to that of an oscillator but determined in frequency by the input only. We have seen the great value of this in the stabilization of transmitters.

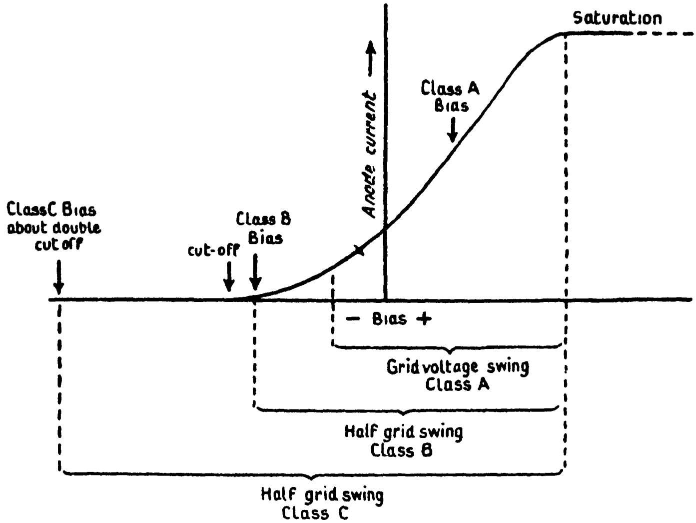


FIG. 87.

The relative operating conditions of classes A, B, and C are shown in Fig. 87, which indicates the appropriate bias, grid input amplitude, and the part of the characteristic utilized. This curve is typical of a valve designed for class-B working, being displaced to the right so that less bias is required. Such a valve would not actually be used for class-A amplification, as it would need positive bias, which we have seen is not permissible. Valves are sometimes designed to work at the class-B

point with zero added bias, for convenience. It will be interesting to compare the efficiency or relative power output obtainable in the three cases, which can be defined as the ratio of radio-frequency power at the input frequency present in the anode circuit to the power drawn from the high-tension source. This can be expressed as $\left(\frac{\text{volts}\times\text{amperes output}}{\text{watts input}}\right)$. The theoretical efficiency of class-A operation cannot exceed about 50 per cent., 40 per cent. being a typical figure. It can be shown that for class B the efficiency of the same valve rises to between 65 and 70 per cent., whilst in the class-C condition it may reach 80 per cent. The exact figures depend upon how near to anode-current saturation the valve is allowed to go. The actual power output for a given input will decrease as a valve is changed from class A to C, because although the efficiency rises the actual anode input power (the mean anode current) falls. If I be the peak anode current, then for class A the mean current will be $I/2$, for exact class-B adjustment it becomes I/π, and for class C it becomes more nearly $I/6$. However, if the input be increased up to the safe limit, a given valve will deliver most power under class-C conditions, provided that the cathode emission is sufficient to yield a very high saturation anode current. For this reason special valves designed for class-C operation in transmitters are provided with exceptionally large cathode emission, and it may be impossible for the anode to withstand this large emission current were the valve to be used in the conventional class-A condition.

We noted in Chapter VI that the self-oscillator has ceased to be widely used for transmission, except at ultra-high frequencies. When used, however, the influence of grid bias upon efficiency is very similar to that explained above for amplifiers, increasing steadily from the class-A to the class-C condition. The self-oscillator can be regarded as an amplifier providing its own grid excitation by reaction, and similar conditions apply to the anode-circuit efficiency. As a rule, bias is derived through the flow of grid current through a leak. Hence the working bias point depends upon the resistance of this leak and upon the grid current, the latter depending upon the reaction coupling and any factors affecting the amplitude of oscillation at the grid. Its calculation is therefore not simple. It may be taken, however, that the self-oscillator will be biased

considerably negative, if it is to work at high efficiency and to convert a large proportion of the anode input power into oscillatory power.

EXAMINATION QUESTIONS

1. What are the essential differences between high-frequency and low-frequency amplifiers? Give diagrams of a two-valve amplifier of each type.

City and Guilds of London Institute. Preliminary Exam. 1937.

2. Why is a large step-up not possible in a high-frequency transformer used as an intervalve coupling? Upon what factors does the step-up depend in this case?

Institute of Wireless Technology. June 1934.

3. A tuned-anode circuit of inductance 300 μH., capacity 0·0001 μF., resistance 10 ohms at the resonant frequency, is coupled to a valve of A.C. resistance 500,000 ohms and amplification factor 1,000. Calculate the P.D. developed across the condenser of the tuned circuit with an input to the grid of the valve of 0·01 volt at the resonant frequency. *I. W. T. November 1934.*

4. Explain why an impedance is essential in the anode circuit of a valve before it can be used as an amplifier. What forms can this impedance take, and what are the characteristics of the amplifier which results?

5. Compare the properties of class A, B, and C amplification. What are the circumstances under which each of these can advantageously be used?

6. In an amplifier employing leak and condenser coupling the value of the leak is 100,000 ohms. If the coupling efficiency is to be 90 per cent. at 50 cycles per second, calculate the value of coupling condenser required. How long will it take a charge on the condenser to leak away to 1/10 of its initial value?

A. M. I. W. T. June 1937.

7. A valve amplifier consists of a valve having an internal resistance of 30,000 ohms and a mutual conductance of 1 mA. per volt. If an external resistance is placed in the anode circuit, calculate its value for a voltage amplification of 10. *A. I. W. T. June 1937.*

8. Describe the action of a three-electrode valve amplifier. A valve has an anode slope resistance of 20,000 ohms and an amplification factor of 16. It is used as a low-frequency amplifier with an anode load of 2 henries inductance and negligible resistance. Calculate the magnification when the input voltage has a frequency of 4 kilocycles per second. *Grad. I. E. E. 1936.*

9. Describe the push-pull method of amplification, and enumerate and explain its advantages over the single-valve method. (See also Chapter XII.) *Grad. I. E. E. 1935.*

10. The voltage amplification given by a three-electrode valve is 3·65 when the load resistance is 15,000 ohms and 5·10 when the load resistance is 30,000 ohms. Find the amplification factor and internal resistance of the valve, proving any formula used.

Grad. I. E. E. 1935.

CHAPTER IX

MODULATION AND RADIO-TELEPHONY

THE variations of air pressure which we know as sound-waves are of a very complex character, and when analysed those corresponding to the simplest word, or even vowel, are found to have very complicated wave forms. There are many ways in which a visible record of the form of the air vibration corresponding to a given sound can be obtained. In one method, a

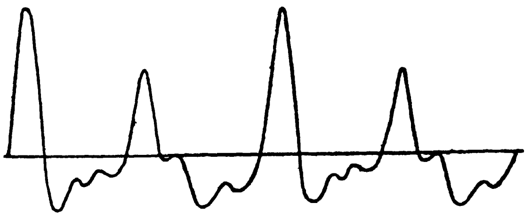


FIG. 88. Wave Form of *a* as in *father*.

gramophone or phonograph record is played slowly, and the motion of the needle magnified by a lever action and recorded. The wave forms corresponding to many different sounds have been studied in this way. Fig. 88, for example, represents the form corresponding to *a* in *father*. The problem of telephony is accurately to reproduce electrically at the distant receiving station the complex sound-waves communicated to the transmitter. When we speak into an ordinary transmitting telephone, the diaphragm of the transmitting microphone is set in motion, and variations of the electric current produced are conveyed along the connecting line which cause the telephone at the receiving end to vibrate in such a way as to reproduce the vibrations of the transmitting microphone. The movements of the receiving diaphragm set up sound-waves in the air corresponding to the sound-waves falling on the transmitting microphone. The sounds spoken into the transmitter are thus communicated to any one listening at the other end.

The simplest form of telephone transmitter is that known as the carbon microphone. Its action is based on the fact that the resistance of a collection of carbon particles depends on the pressure exerted on them. Fig. 89 shows a common type of

such a microphone known as the 'solid-back' transmitter. *D* is the diaphragm upon which the sound-waves fall. *T* is the solid back, on which a metal cup *B* is mounted containing the carbon granules *C*. The button *L* is maintained in contact with the diaphragm by the spring *S*. *E* and *F* are two hard carbon plates closing the cup. These are connected to the outside circuit in

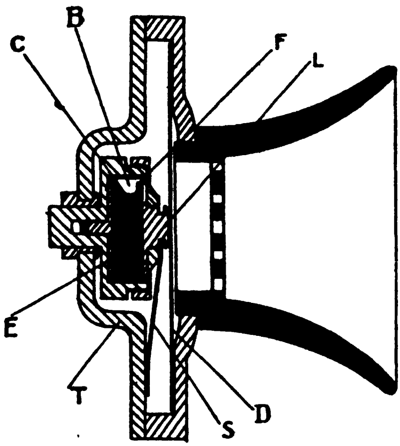


FIG. 89. Carbon Microphone.

series with a battery of low voltage, and are insulated so that any current from the battery in passing from *E* to *F* must pass through the entire mass of carbon particles. The vibrations produced by the voice or other sounds set the diaphragm in motion, and the movements of the diaphragm are communicated to the electrode *F*, so that the pressure on the carbon granules is varied, and changes of current corresponding to their variations in resistance flow in the 'line' circuit. The spring *S* prevents the vibrations which the diaphragm might make at its own low frequency, that is to say, it provides one method of resisting and damping out free vibrations. In many modern microphones the diaphragm is so tightly stretched that its natural frequency is above the main audible range, and the spring *S* is unnecessary.

This simple form of carbon microphone is by no means a perfect device for the conversion of sound-waves into current variations, as it is not equally responsive to all acoustic frequencies and, through resonances in the diaphragm and air

cavities, imparts a coloration of its own to the sounds. We shall compare the distortion produced by various types of microphone when discussing quality of reproduction in a later chapter, but for the moment it will be sufficient to review the other methods by which sound-waves can be transformed into electrical variations.

The carbon microphone illustrated is termed a solid-back single-button type, and has been used for years in the ordinary Post Office telephone service. It is cheap and delivers clearly intelligible speech. It can be improved by the use of two buttons, the second of which is placed opposite to the first on the other

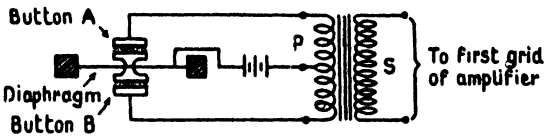


FIG. 90. Double Button Microphone.

side of the diaphragm. The two buttons are compressed by the vibrating diaphragm alternately, in a form of 'push-pull' operation, so that an increase in current through the one corresponds to a decrease in the other. The fluctuating currents from the two are passed through two windings of a transformer in opposite directions, so that they produce an additive output from a common secondary winding. Not only is the output from the microphone increased by this arrangement, illustrated in Fig. 90, but there is a tendency for spurious harmonics of the diaphragm vibration to cancel in the output circuit. In the interests of good quality the diaphragm should be tightly stretched and well damped. This reduces free or resonant vibration, but also limits the amplitude of vibration set up by the sound-waves, rendering the microphone less sensitive. In spite of this, however, the carbon microphone is the most sensitive type in common use.

A second improved type is the 'transverse-current' carbon microphone, in which the current flows through the carbon granules in a direction parallel to the diaphragm, between pairs of paralleled electrodes immersed in the granule-container. The diaphragm forms one side of this container and in vibrating varies the pressure between the granules and consequently the resistance between the pairs of electrodes. An advantage of this construction is that the diaphragm forms no part of the electrical

circuit, and can be of less mass and of non-conducting material such as mica. Also the change of resistance for a given amplitude of the diaphragm will be greater, because of the increased area of electrode surface and volume of granule-filled space. Carbon electrodes are used, the only other material which gives an equally low noise-level being a gold-plated surface. The latter is suitable for fixed electrodes or a conducting diaphragm when in direct contact with the granules.

Whilst long the most popular, the carbon microphone suffers from certain defects. A degree of background noise caused by imperfect contact between the granules is inherent and gives rise to a hiss which becomes serious if the microphone is followed by considerable amplification to pick up distant sounds. The need for an exciting current is often inconvenient, whilst the quality from carbon microphones is not altogether satisfactory for musical reproduction or the exacting purposes of modern broadcasting.

To overcome these difficulties a number of microphones have been designed in which a potential more strictly proportional to the pressure or velocity of sound-waves upon the diaphragm is directly generated. These types need no exciting current, but act as direct transformers of acoustical into electrical energy. They contain no variable contacts and are entirely free from self-generated noise. Naturally, however, the output potentials are small, since they are directly obtained from the very small energy of the incident sounds; and these microphones have only become useful with the help of modern high-amplification technique. They are characterized, however, by small and light moving parts, the low inertia of which enables a good response free from resonances to be obtained over the whole range of audible frequencies.

Microphones can be classified according to the electrical property upon which their action depends, the carbon microphone being a 'resistance-variation' type, or 'resistance microphone'. One of the earliest improved types was the 'condenser microphone', which consists of a rigid metal back-plate, parallel to which and very close to it is a metal diaphragm. The latter is very thin and tightly stretched, so that its natural period of vibration is above the highest useful musical frequency. These two electrodes form a condenser, and as the diaphragm vibrates the capacity of the condenser will vary. If the condenser is

charged with a constant quantity of electricity by the application of a high potential to the electrodes through a choke or high resistance, then the potential across its plates will vary inversely as any small capacity change. For since $Q = E/C$, and if Q be maintained constant, E must vary inversely with C, no matter at what frequency C may vary. These potential fluctuations correspond very closely to the sound-waves reaching the diaphragm, reproduction being good, but a practical drawback lies in the need for a high polarizing potential, which demands very careful insulation. The potential variations between the electrodes are transferred directly to the grid of the first amplifying valve, usually by earthing the back-plate and by coupling the diaphragm through a condenser, as shown in Fig. 91. As the reactance of the microphone is very high, a high grid leak must be used, 10 megohms being usual. Also as the output is so small very careful screening against outside electrical disturbances is necessary. It is usual to keep the grid lead very short by enclosing the first amplifying stages in a common metal case with the microphone.

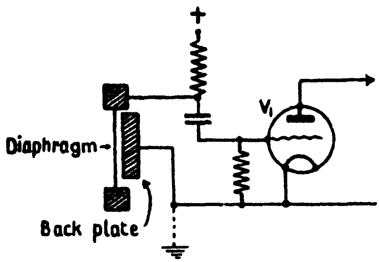


FIG. 91. Condenser Microphone.

A third class of microphones employs the potentials induced in a moving conductor immersed in a strong transverse magnetic field. The moving-coil type is similar to the well-known loud speakers in construction. The diaphragm may be a light cone, to which is attached a light coil of a few turns of wire free to vibrate in the gap of a strong cobalt-iron magnetic system. The ribbon microphone employs a single light metal ribbon which is free to vibrate across the intense field between the poles of a permanent magnet. The ribbon is itself the 'diaphragm', and of extremely small inertia. Both these microphones can be described under the general term of 'magnetophone' and are analogous to the dynamo in action. The moving conductor is of low reactance, and the currents induced in it are at very low potential, hence a step-up transformer is used close to the microphone, increasing the output potentials to a magnitude more suited to the grid circuit of a valve.

Yet a fourth type of microphone now increasing in favour is the piezo-electric type, in which the potential variations set up by varying pressure suitably applied to a crystal, as explained in Chapter VI, are made use of. The crystals employed are of Rochelle salt, which has the highest piezo-electric constant available, and are protected from damp air by coatings of wax or rubber. The crystal itself may form the 'diaphragm', in which case a bank of several in series-parallel is used to increase the output; or a single crystal 'cell' may be driven by any conventional form of diaphragm or cone. The output of crystal microphones is somewhat greater than the preceding types, and the quality can be almost as good. We shall discuss the directional and reproducing properties of microphones in a later chapter, however, and, having noted the methods by which they convert sound-waves into corresponding potential variations, continue now with the discussion of modulation in radio-communication.

In radio-telephony, in place of the current in the connecting wires in land-telephony, the amplitude of a pure continuous sine wave of any convenient frequency emitted from the transmitter is varied periodically by the sounds. The continuous wave, which is called the 'carrier wave', is thereby moulded into a form corresponding exactly to the shape of the sound-wave.

This process is known as modulation. Thus Fig. 88 represents the wave form of currents in a microphone corresponding to the sound *a*, while Fig. 92 represents a continuous carrier wave modulated by the sound-wave. This is the form in which the sound *a* would be radiated into space from a radio-telephony transmitter of the most usual type.

If a suitable detector is used at the receiving end, the oscillatory current corresponding to Fig. 92 can be rectified into a direct current and the sound produced in the receiving telephone will correspond to the original speech.

The amplitude of the modulated continuous wave radiated varies periodically above and below certain positive and negative values which are not zero. In Fig. 92 these values are represented by dotted horizontal lines.

The essentials of a radio-telephony transmitter are clearly a means of generating a pure and steady continuous-wave oscillation, which we have studied in Chapter VI, and a means of

varying the amplitude or modulating this wave faithfully by the current or potential variations produced by a microphone. Any source of continuous waves will serve for telephony purposes, but it is best that it should be of good frequency stability, and produced from a crystal or master oscillator and multiplying stages in the manner that has been described. This requirement

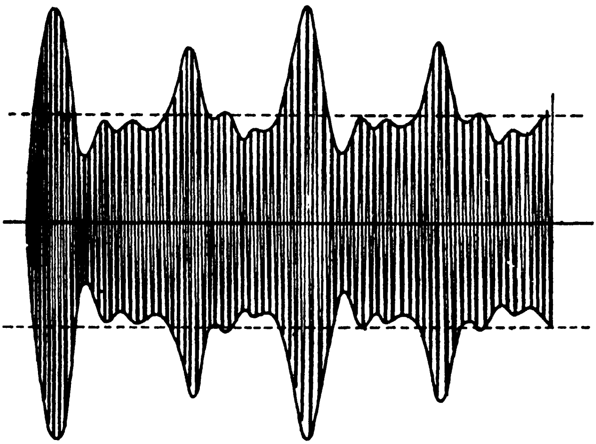


FIG. 92. Continuous Carrier Wave modulated by Letter *a*.

is even more necessary than for telegraphic signalling, because an unstabilized carrier wave may vary in frequency as a result of the modulation process, with resulting distortion in a selective receiver as the carrier wave fluctuates in and out of tune with the latter. Self-excited oscillators can be used for telephony, and are so used in portable equipment, where bulk and weight are important, and also at ultra-short waves, where selectivity may not be very great; but they find no place in modern commercial point-to-point services or in broadcasting.

Referring to the illustration of a modulated wave in Fig. 92, the dotted lines show the amplitude of the original carrier wave before modulation. Clearly if the peaks of modulation reduce the carrier momentarily to zero, on the one hand, whilst doubling its amplitude on the other, a condition of maximum or 100-per-cent. modulation exists. Any deeper modulation will cause distortion, because the peaks of the carrier envelope will be cut off, and on detection the wave form produced will differ from

that which left the microphone. Percentage modulation is an important conception and can be defined as the ratio of the maximum increase in carrier amplitude to the mean carrier amplitude. In practice it is usual to limit the maximum modulation to some 80 per cent., the mean value lying between 20 and 30 per cent. only.

It is interesting to consider the effect of modulation upon the radiated power. The power of a steady carrier can be measured by the square of the carrier amplitude, either root mean square mean potential or current, because, on the assumption that the circuit impedance remains constant, the power will be proportional to current squared. Thus the power without modulation is given by I^2, where I is the mean carrier current. For the case of 100-per-cent. modulation the minimum carrier power will be zero, whilst at the maximum peaks the current becomes $2I$, and the power $(2I)^2 = 4I^2$. This extra power can only come from the modulation circuits, and therefore it is clear that additional power must be supplied to effect modulation of an oscillator. This fact can only be overcome if the steady carrier be reduced and then modulated in such a way that its peak value never exceeds the maximum output obtainable from the oscillator. Modulation still corresponds to increased peak power, but this is now obtained from the oscillator itself, though only at the expense of a carrier energy one-quarter of that possible in telegraphic working.

It is thus possible to distinguish two basic forms of modulation in practice, which may be termed 'additive' and 'subtractive' modulation. This is not a generally used distinction, but is very helpful in showing how modulating circuits operate. An example of additive modulation would be a simple valve oscillator in which the amplitude is varied by the alteration of the high-tension anode voltage above or below its normal value. This could be done by adding in series with the anode supply circuit the output voltages of a large low-frequency modulation-amplifier delivering considerable power. Thus on speech peaks the anode voltage of the oscillator will be increased by the peak voltage of the amplified modulation, and the carrier power will be increased. This is an example of the system termed 'anode modulation'.

In the case of 100-per-cent. additive modulation it can be shown, by integrating the expression for the instantaneous

power over a time period corresponding to one cycle of a sine-wave modulating note, that the mean radio-frequency power is increased 1·5 times. The mean carrier current, in an aerial, for example, will be increased by the square root of this, approximately 1·23 times. Consequently, if the efficiency of the transmitter remains unchanged, the total power supplied to the anode circuits must increase at least 1·5 times on full modulation. This is characteristic of additive-modulation circuits, the extra power coming from the modulator.

Subtractive modulation would be obtained in the simple example of a self-excited oscillator, if a microphone were inserted into the aerial circuit so that the resistance of the microphone varied the actual radio-frequency aerial current. This was a method used in some early transmitters, banks of carbon microphones being used in parallel to handle the large current without overheating. In this case the aerial current would be reduced to half by suitable selection of microphone resistance, and the oscillator would continue to run at full efficiency. On speaking into the microphones the aerial current would be directly varied, and its mean value would rise 1·23 times on full modulation. This extra power would come from the anode circuit of the oscillator in the form of increased anode current, and not from the modulating circuits. A similar form of modulation would occur if the oscillator power output were reduced to one-quarter maximum by the application of high negative grid bias. If the amplified modulation potentials be now superimposed upon the grid bias, the output from the oscillator would be modulated, its peak value being equal to the original unbiased output. This is termed 'grid-bias modulation', and again the power required to produce the peak carrier amplitude comes from the oscillator anode circuit itself, and not from the modulating circuits, which merely act as a controlling factor. Clearly, if these peaks are not to overload the oscillator, the mean carrier power must be cut down to one-quarter maximum before modulation is applied and the rating of the transmitter is reduced. Additive modulation is therefore the most economical, for it enables a transmitter of normal carrier power P to deliver a mean telephony power output of $1{\cdot}5P$, with peak power of $4P$; whereas with a subtractive system the same transmitter can only deliver a mean power on telephony of $1{\cdot}5P/4$, and a peak only of P.

The various forms of modulating circuits in practical use must now be considered. They are:

(*a*) Anode-circuit modulation.

(*b*) Absorption modulation.

(*c*) Grid-bias modulation.

The characteristic of the first of these is essentially additive, whilst the other two are more subtractive in nature. The essential principle involved in anode modulation is the amplification of the speech potentials from a microphone until considerable power is available at speech frequencies. This power is then added to the anode supply current of the transmitting valve to be modulated, which may be either an oscillator or an amplifying stage forming part of a crystal-controlled transmitter.

There are three methods by which this amplified speech power can be added to the anode supply. They are: firstly, parallel connexion, which has been known from the earliest days as 'choke control'; secondly, series connexion, now becoming increasingly popular under the name 'series modulation'; and transformer coupling.

A simple form of choke control is shown in Fig. 93. Here the microphone potentials are shown amplified by one triode stage and then fed to the grid of three valves wired in parallel to form a single modulating stage of large current-carrying capacity, V_2, V_3, V_4. It will be seen that the anode circuit of this modulating stage is fed with high-tension current in parallel with that of the transmitting valve V_5 through the common choke marked 'speech choke', which gives the circuit its old name. V_5 is shown as a simple self-oscillator, and it is true that this circuit is one of the best possible for the modulation of such a simple transmitter; but V_5 may equally well be regarded as the final power-amplifier stage of a driven transmitter. The arrangement shown is typical of the earlier telephone transmitters in common use about the year 1927.

The operation of the circuit is simple. The modulating stage can be regarded as a power amplifier (see Chapter XII) which is choke-coupled to the anode circuit of V_5 and thus adds its power output to the anode current normally reaching that valve through the choke. As the potential of the anodes of the modulating valves rises and falls, these changes are communicated to the anode of V_5, and the radio-frequency output of this valve will vary accordingly. The air-cored choke shown in the

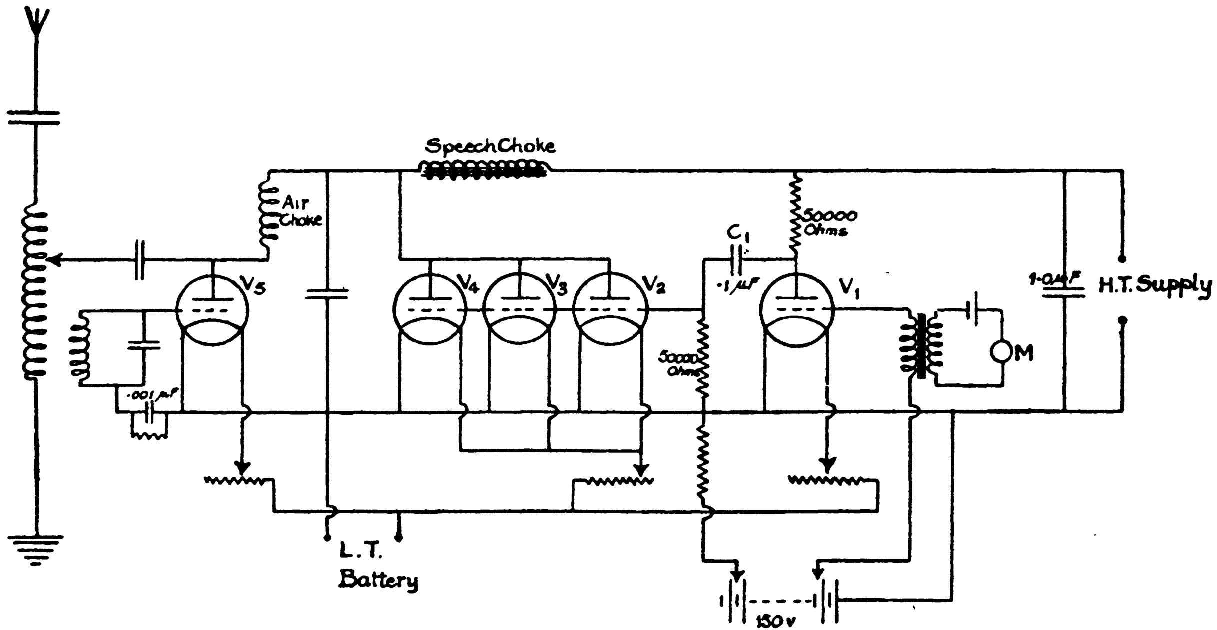


Fig. 93. Choke-control Radio-telephony Transmitter.

diagram is necessary to prevent radio-frequency current from passing back into the modulator, but plays no other part in the operation, its reactance to speech frequencies being negligible.

The circuit can be explained in an entirely different way also, and was originally designed from this viewpoint. The speech choke is iron-cored and of high inductance. It therefore offers a high reactance to the speech frequencies and practically prevents variations in the current through it. Thus the total anode current of the modulator stage and the modulated stage V_5 taken together will be maintained constant while modulation goes on, and if one falls the other must rise proportionately. Hence, as the anode current of the modulators varies with the amplified microphone potentials, identical but inverted changes must occur in the anode current of V_5. We have shown that as the anode current of this stage varies, so will the radio-frequency carrier output. It is true this will be in opposite phase to that of the modulator anode currents, but since all speech frequencies are equally reversed in phase there is no relative change, and distortion does not result.

It will be noticed that three modulator valves are shown in parallel. This was done to stress the fact that for full modulation it is necessary for the modulator to handle about four times the power that the radio-frequency stage handles. Thus if similar valves are used for both oscillator and modulators, the former being fully loaded, then three or four modulators in parallel will be necessary to handle the necessary power. At one time banks of valves in parallel as shown were often employed, but to-day it is usually possible to obtain a single valve having four times the power-handling capacity of the valve V_5 and which will suffice alone. Regarded from the power viewpoint it is simple to see the need for this large modulating stage. At peaks of 100-per-cent. modulation it has been shown that the carrier power rises momentarily four times. Since this power comes from the modulator stage, clearly at least three times the power will momentarily be drawn from this stage. If all valves are biased to work on the linear portions of their characteristics, little opportunity for distortion occurs in choke modulation. Quality is excellent, and for many years it has been the most favoured circuit for broadcast transmission and high-quality telephony. There is, however, one drawback, namely lack of economy. The modulator valves draw a con-

siderable steady anode current, and working in class A they can only deliver as speech energy to the transmitter about 20 per cent. of the power which this represents. This means in practice that the power supplied to the modulator stage exceeds that to the transmitter several-fold, and, except during extreme modulation peaks, most of this power is wasted as heat at the valve anodes.

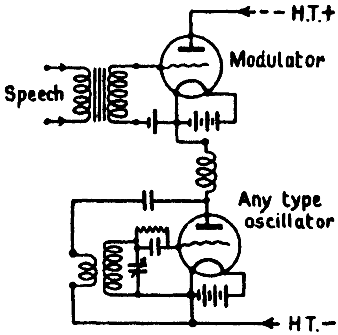


FIG. 94. Series Modulation.

Mainly for this reason series modulation has come into use. The circuit is similar except that the modulator valve is in *series* with the anode supply to the transmitter (Fig. 94). This means, of course, that the cathode of the one, usually the radio-frequency stage V_5, will be at high potential above earth, and must be carefully insulated and fed from a separate filament supply. Now since the valves are in series the same current must flow through both, and as this varies with the microphone potentials applied to the modulator, identical changes must occur in the anode current of V_5. The high-tension voltage must be raised in this system to the sum of that required for both valves taken separately, but since the same current flows through both there is no increase in current requirements. The total power will now be approximately doubled, owing to the increased voltage, but will not be increased several-fold as in the case of choke control. Thus the system is more economical, whilst retaining the good-quality characteristics of the older system, and for these reasons is adopted in the most recent broadcasting stations. Series modulation can be understood more easily if the modulator be regarded simply as a resistance connected in the high-tension supply lead to the transmitter, but having the special property of changing its value with the microphone potentials. It is by treating the modulator valve as a resistance that suitable operating values are calculated in practice.

The third form of anode modulation is a little less simple and not so easily designed to eliminate distortion. In fact it is doubtful if it can be made so distortionless as the series system in

such exacting cases as television transmission, but it has the advantage of even greater economy of power. We have seen that the oscillatory power obtainable from an amplifying stage is much greater if the valve be operated in class B or C, but that push-pull working is essential if distortion is to be minimized. In a later chapter it will be shown that very much more

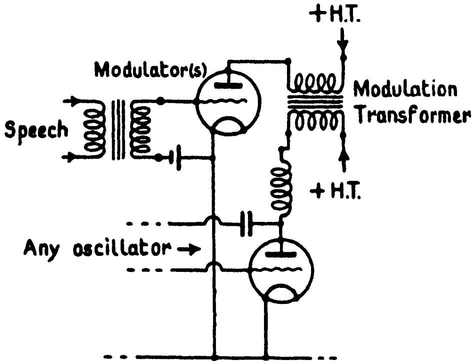


FIG. 95. Injected Modulation.

power is obtainable from the output stage of a speech-amplifier for a given high-tension consumption if a pair of valves are used in this way. Neither the choke-control nor series-modulation system lends itself to the use of push-pull modulators or to any but class-A biasing of the modulators. In the present system the modulation-amplifier is terminated in a power stage of the most efficient class B type, exactly similar to those we shall study later when discussing loud-speaker reproduction, and the amplified speech energy is injected into the transmitter anode circuit through a transformer (Fig. 95). The primary of this receives the power from the anodes of the modulator valves, probably in push-pull, whilst the secondary is in series with the normal steady anode supply to the transmitter stage, to which it adds the modulating power. The transformer ratio is selected so that the primary matches the impedance of the amplifier output stage, whilst the secondary matches that of the transmitter anode circuit. In this way the maximum transfer of energy is obtained between the two.

It is not possible to devote further space to anode modulation, but it should be stressed that it is the most widely used and generally most satisfactory system. This is partly because of the

increase in radiated energy which it produces and which is not obtainable from the other systems. On the other hand, subtractive modulation does not need large-power amplifying stages consuming heavy currents, and is in general cheaper, less complicated, and better suited to lower-powered or mobile equipment. Absorption modulation has played an important part in early radio-engineering, but is now less used, with the possible exception of ultra-short-wave transmitters working near the upper limit of frequency. The simple case quoted, of a microphone used in the aerial circuit, is typical of absorption modulation, although long obsolete. The basic principle is to operate the transmitter at full output and to remove a part of the radio-frequency energy from the aerial circuit by the absorbing modulating device. This might be compared to a tap by which energy is fed to the aerial at will, under the control of the speech potentials. Another early form of absorption modulation utilized a series circuit of inductance, capacity, and resistance, loosely coupled to the transmitter aerial coil, and tuned near to the radiated frequency. Such a closed circuit absorbs energy from the aerial, and if the resistance be a carbon microphone, the energy drawn off will vary as the microphone is spoken into. At one time many elaborate microphones were designed for the special purpose of handling moderately large radio-frequency power without overheating.

An ingenious application of absorption modulation has been made to micro-wave transmission, where normal methods of control were very difficult to apply. The transmitter radiated a narrow beam from an aerial at the focal point of a parabolic reflecting system. In the path of this beam was placed a long glass tube bent back and forth upon itself to resemble a large plane screen. This contained ionized gas at low pressure, being similar to the familiar neon signs in construction. The ionized gas, being partially conductive, absorbed the passing radiation to an extent depending upon the intensity of ionization which was produced by amplified speech potentials controlling an oscillatory ionizing 'carrier' through the tube. In this way the radiated beam was directly controlled and good-quality modulation obtained.

A form of absorption more recently used, and still found in certain transmitters, may be termed valve absorption. Here a valve is used to form a variable damping resistance across part

of the aerial circuit or a circuit coupled to it. The principle is shown diagrammatically in Fig. 96. The effect of an increase of grid potential in a valve is to reduce the apparent internal resistance of the valve through increase of anode current, while similarly a decrease of grid potential increases the internal resistance. A portion of the aerial coil L of the transmitter is

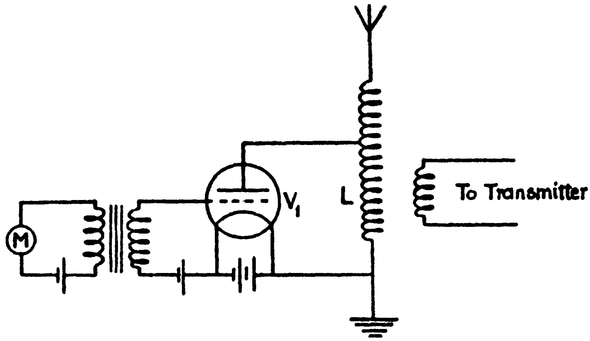


FIG. 96.

connected between the anode and filament of the valve V_1. This is equivalent to connecting across L a resistance which will vary as the grid potential of V_1 is varied by the speech currents produced by the microphone. Thus, as M is spoken into, the damping of the aerial circuit is varied, and the carrier wave is therefore modulated in accordance with the speech currents.

Grid-bias modulation is perhaps the simplest possible system, and is therefore used in mobile and low-power equipment more widely than in high-powered stations. In effect, it is subtractive and gives an over-all efficiency not very different from the absorption methods. Unfortunately, however, it is very difficult to adjust the system to a low level of distortion, and although this is possible, the tendency is for quality to fall below either of the preceding systems. This follows from the fact that for modulation by grid-bias variation to occur at all it is essential for the modulated valve to work in a non-linear condition.

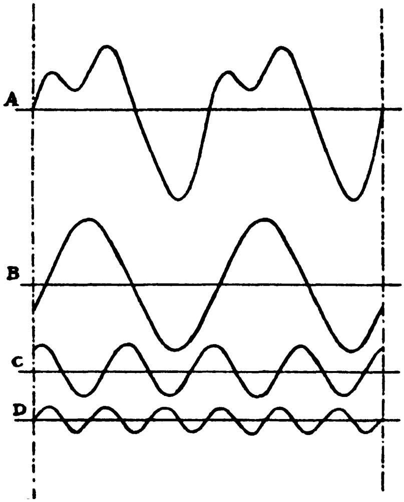
The system can be applied either to a self-oscillating valve or to an amplifying stage. The latter is preferable, because it is only when applied to a radio-frequency amplifying stage biased to the class C condition that freedom from distortion is theoretically possible. The chief advantage of grid modulation is that very little power is drawn from the modulating stage. The

considerable power-loss inherent in the choke or absorption modulator is avoided, and the final stage of the modulation-amplifier need supply only sufficient power to provide any grid current which may flow through the modulated stage.

In operation, the modulated stage is biased from some permanent source to a high negative value, sufficient to reduce the radio-frequency power output of the transmitter to one-quarter of its maximum value. The amplified speech currents are then applied to the grid, usually from the secondary of a low-frequency coupling transformer inserted at the low-potential end of the grid leak or in series with the grid tuning inductance, when it must be by-passed by a suitable condenser of low radio-frequency reactance. The speech potentials then vary the grid bias, and thus the output to the aerial is controlled.

For high-power telephony over long ranges, the arrangements hitherto described possess certain disadvantages. For the proper understanding of these and of the methods used to overcome them it will be necessary to consider in a little more detail what the wave form of a sound-wave due to speech or music really represents. It was stated on page 26 that an oscillation of any form could be considered as composed of a number of different pure sine waves of various frequencies. Now, a complicated speech wave is made up of a number of component sine waves corresponding to a large number of frequencies ranging between 100 and 3,000 cycles per second, while, with music, frequencies as high as 15,000 cycles per second may be present. For example, the three sine waves B, C, D making up the violin note A are shown in Fig. 97. Similarly, the more complicated modulated wave given in Fig. 92 is the result of modulating the carrier wave simultaneously with a number of sine waves of various frequencies. If we now suppose, in order to simplify the problem, that instead of modulating the carrier wave, whose frequency we will take to be f, with a complicated form of sound-wave, we modulate it with a single simple sine wave of frequency p, as shown in Fig. 98, a comparatively simple mathematical analysis would show that the modulated carrier wave can be considered at any moment to be composed of three waves whose frequencies are f, $f+p$, and $f-p$. The production of these new frequencies is explained in the 'sideband theory' which we shall study in Chapter XI. They must be introduced here in order to explain

the more advanced ideas on modulation which follow. In the systems of telephony which we have considered so far these waves heterodyne in an ordinary receiver so as to produce a beat note of frequency p, i.e. a sound corresponding to that exciting the transmitter. Now it is clear, from this way



Curve of violin note and its harmonic components

FIG. 97.

of looking at the matter, that the carrier wave (frequency f) has no effect on the transmission of the sound except to provide a wave which will be of use in heterodyning the waves of frequencies $f+p$ and $f-p$, which really convey the sound we are wishing to transmit.

A considerable amount of power is used to transmit the carrier wave; and it will be obvious that in long-distance working great economy can be achieved by not expending power in radiating the carrier frequency f from the aerial, but employing instead a small local oscillator of negligible output to supply the heterodyning frequency f at the receiving end. This locally generated oscillation will produce the desired frequency p in the receiver by heterodyning with either of the frequencies $f+p$ or $f-p$. Hence, not only is the carrier wave unnecessary, but one or other of the component frequencies is also unnecessary. Thus, if an arrangement is used by which the wave (say) of frequency

$f+p$ only is transmitted, and this frequency is heterodyned at the other end by a frequency f, good speech will be obtained in the receiver.

Not only will there be a saving of about 75 per cent. in power by transmitting one only of these 'sidebands', as the waves corresponding to the frequencies $f+p$ and $f-p$ are called, but there is also another great advantage in that the interference with other lines of communication is reduced. This interference

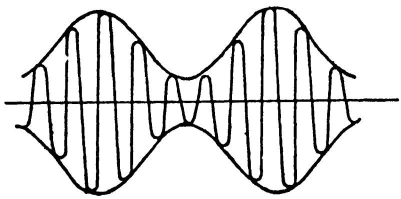


FIG. 98. Carrier Wave modulated with a simple Sine Wave.

factor is a very serious matter in the development of high-power telephony.

Now, the wave forms corresponding to speech may contain as a minimum the frequencies between 100 and 3,000 cycles per second. Thus, if we transmit in the usual manner a carrier wave and two sidebands, each of these sidebands will cover 3,000 cycles, making a total of 6,000 cycles occupied by the transmission. For example, if the carrier wave were of 20,000 frequency, then an ordinary radio-telephony transmitter would emit all waves of frequencies from 17,000 to 23,000 cycles per second. It is clear that on these long waves only a few lines of radio-telephony will be possible without the various frequencies emitted overlapping and causing interference. Even then, no allowance has been made for the various radio-telegraph services taking place on long waves. Manifestly, therefore, it is very important, if possible, to adopt some method requiring the use of fewer cycles. If only one sideband is used, covering a range of 3,000 cycles, the number of different radio-telephonic communications possible is doubled. Or, alternatively, a considerable number of additional radio-telegraphic communications will be possible, since about ten radio-telegraphic lines of communication can be got into the range of frequency required for one radio-telephony communication.

The arrangement necessary in practice for separating out the

frequencies, so that only one sideband is radiated, is somewhat complicated. That used in certain of the earliest experiments has been described to the American Institute of Electrical Engineers.* The principle of the apparatus employed is shown in Fig. 99. The apparatus consists of two balanced three-electrode valves connected as shown in the figure. (For simplification in drawing,

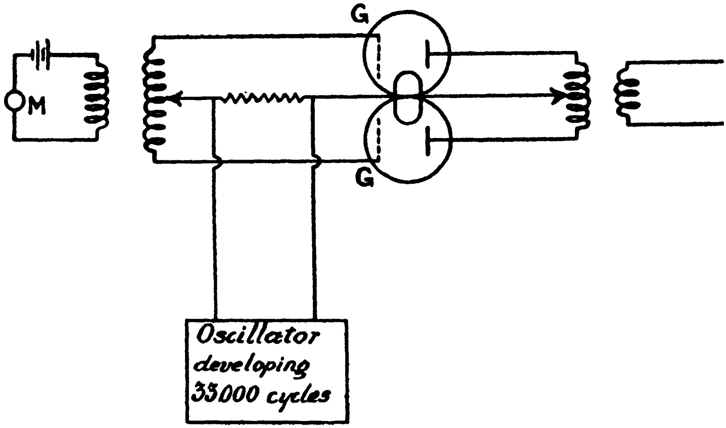


Fig. 99.

the filament is shown between the grid and the plate, and the filament batteries are omitted.) A carrier frequency of 33,000 cycles is developed by means of a valve generator and applied at the middle of the system as shown. When the microphone is not being spoken into, equal and opposite potentials will be applied to the grids (G) of the valves, and hence equal and opposite currents will flow in the two halves of the transformer connected to the anodes of the valves. Thus the carrier frequency is balanced out in the system. When the microphone is spoken into, however, the balance is destroyed, and the system will deliver to the secondary of the anode transformer the two sidebands containing frequencies from 30,000 to 32,700 cycles and from 33,300 to 36,000 cycles respectively. These oscillations are then passed through a filter circuit, which is composed of acceptor and rejector circuits so adjusted as only to pass frequencies between 30,000 and 32,700 cycles. These frequencies then pass through a second pair of balanced valves, which are supplied by an oscillator generating a frequency of 88,500 cycles.

* For abstract see *Electrician*, vol. 91, 1923, p. 424.

The frequencies developed in the anode transformer of this balanced modulator are thereby further separated and are applied to a filter circuit passing frequencies of from 55,800 to 58,500 cycles. These frequencies are then passed through a series of three amplifiers, until the small power, viz. 5 watts, developed in the first pair of balanced valves is amplified in stages up to 150 Kw., which is the power applied to the aerial. The band of frequencies emitted from the aerial is then from 55,800 to 58,500 cycles.

In the receiving system it is necessary to remodulate the side-band radiated from the transmitter with the corresponding carrier-wave frequency which has not been transmitted, i.e. a wave of frequency of 55,500 cycles. When the emitted wave is heterodyned by oscillations of this frequency, it is clear that frequencies included between 300 and 3,000 cycles corresponding to the voice frequencies will result, which can then be detected in the usual manner.

It will be seen that transmission with this system of side-band telephony is secret to the extent that no intelligible speech can be received unless the incoming oscillations are heterodyned with oscillations from a local generator carefully adjusted to the correct frequency of the carrier wave. Nothing would be heard in an ordinary crystal receiver, for example, or in any receiver suitable for broadcast reception. Hence the system, while well adapted to a commercial service from point-to-point working, is not suitable for broadcasting. The chief practical difficulty arising in its use is that of the local oscillator which reinserts the carrier at the receiver. Not only must this be of the correct frequency and amplitude, but it must also be maintained in the correct phase. The latter requirement gave so much difficulty in practice that it delayed the general use of single sideband working for many years. It can be overcome most simply by the radiation of a small percentage of the original carrier, which can be amplified at the receiver before recombination with the incoming sideband frequencies. The quality of reproduction by this system is adequate for telegraphic working and for intelligible speech, for which it is now widely used, but it is not adequate in the present state of development for the first-class reproduction of music, picture-telegraphy, or television.

Clearly single sideband working provides a degree of secrecy not obtainable from the symmetrical type of transmission, but

where secret communication is seriously desired it remains inadequate because anybody using a local oscillator together with a normal receiver can reproduce intelligible speech. For the long-distance radio-telephone services of to-day other methods of obtaining secrecy have been evolved. An early successful method employed the reversal of speech frequencies at the transmitter, such that high frequencies became proportionately low, and the low became high. This can be done by the use of a moderately low-frequency oscillator, against which the speech frequencies are allowed to beat, producing two sidebands. One of these is inverted with respect to the original speech, and this band is filtered off and used to modulate the transmitter. The system is not quite secret, however, since if the original heterodyning frequency be added at the receiver speech becomes recognizable. Also it was found that a self-oscillating detector can reinvert the speech frequencies under certain conditions of adjustment, and that private conversations were being read by such means. An additional degree of secrecy to overcome this can be obtained by cyclically 'wobbling' the transmitted carrier frequency by about half a kilocycle and providing a synchronous wobble of the receiving oscillator. It is very difficult for a receiver not knowing the exact wobble frequency and amplitude to reproduce this condition, except with the help of elaborate equipment.

Radio-communication is by its very nature public, the majority of messages being broadcast to all suitable receivers within range. Some improvement in secrecy has resulted from the use of directional aerial systems, skip-distance effects, and unspecified short wavelengths, but it is still necessary to employ elaborate equipment if true privacy is to be relied upon. The privacy systems just mentioned have been superseded in many cases by still more effective ones. The underlying principle now used is that of dividing the speech-frequency band into several parts by means of electrical filter networks. These portions are then interchanged, and some may be also inverted, so that the frequency spectrum radiated differs entirely from the original speech. At the receiver similar circuits are used to effect the inversion and recombination of the 'scrambled' speech into its original frequency relationship. It is possible, by prearrangement between the transmitting and receiving stations, to change the sequence of scrambling from time to time, and thus to

make unauthorized listening so difficult as to be virtually impossible.

A review of modulation methods would not be complete without reference to frequency modulation. Up to the present modulation has been described as the variation in amplitude of a constant-frequency carrier wave, the amplitude changes being a copy of the original microphone potentials, whilst steps are taken to keep the carrier frequency as constant as is reasonably possible. Even in this case, however, we have seen that the fact of modulation gives rise fundamentally to a band of radiated frequencies extending on each side of the carrier to a width equal to the highest modulation frequency. The production of these sidebands and their physical significance will be dealt with more fully in the chapter entitled 'Selectivity'.

Consideration will show that since all that is required is to transmit information about a single variable factor, namely the microphone potentials, we can convey this by varying any single factor in the radiated wave. Speaking mathematically, information of the magnitude of a single variable quantity (air pressure at the microphone diaphragm) can be translated into corresponding variations of any one parameter of the transmitted wave, all others remaining constant. The parameter usually chosen is amplitude, but on this reasoning it will be equally possible to maintain the carrier amplitude constant, and to modulate the carrier frequency. An increase in microphone current or potential of, say, twofold might now be represented by a change in carrier frequency of, say, 2 per cent., instead of the usual change of twofold in amplitude. It is just as possible to transmit speech or any other kind of signals in this way.

The modulation of a constant-amplitude carrier by variation of frequency is termed frequency modulation. It has been successfully employed for commercial communication and is capable of excellent results, provided the whole system is designed to make use of it. The presence of partial frequency modulation often occurs as a fault in amplitude-modulated transmissions, particularly when crystal or any similar control is not used. In this case it can only give rise to serious distortion, and the complexity of modern equipment is largely directed to its avoidance. It is quite easy to see that if the transmitted carrier wave is varying in frequency, that is in wavelength, to

correspond to the modulation, such variation can be converted into changes of amplitude in the receiver, and listened to in the ordinary way after detection. This will automatically occur if the receiver be very selective, since as the carrier moves in and out of tune with the receiver the signal from the detector stage must vary with it. To prevent distortion the resonance curve of the receiver should be linear and of suitable slope over the range through which the carrier frequency is varied, and it will be shown later that this can be simply arrived at.

It has been argued that frequency modulation should be superior to the more usual system, because the transmission can occupy as narrow a wave band as is desired, and sidebands need not spread so far on either side of the carrier frequency. Transmissions could thus be placed on closely adjacent wave-lengths, and interference reduced. This theory has been proved to be unsound, and mathematical analysis shows that sidebands are formed in a very similar manner to amplitude modulation. Their width is at least as great, and will in general be much greater. There is thus little reason to prefer frequency modulation, and it has found no wide application. It is claimed to possess certain other advantages, which include more economical use of the radiated energy, less susceptibility to fading, and less responsiveness to atmospherics. It has also been suggested that two independent signals might be modulated upon a single carrier by the use of the two modulation systems simultaneously, and received without mutual interference; or that frequency-modulated transmissions might be inserted with safety between many existing services. There is undoubtedly a degree of truth in these claims, and much thought is being devoted to the problem to-day in the hopes of evolving a generally more satisfactory system.

Up to now modulation has been referred to with reference to the important case of telephony only. There are other forms of intelligence besides speech which can be conveyed over a radio-carrier, and the principles of modulation remain the same for all. Thus picture-telegraphy, which is transmitted in the form of varying current impulses over land lines, can also be handled over a radio-channel. Television follows naturally, being equivalent to the former very much speeded up. We shall study these forms of modulation in later chapters. Meanwhile, telegraphic signalling can also be treated as a form of modulation, and in the

case of modern high signalling speeds, of from 200 to 300 words per minute, no other treatment is suitable. Hand sending of the Morse code at low speeds has been described as the breaking-up of a steady carrier wave into long and short periods, seemingly quite unlike telephony modulation. Strictly, however, it is a special case of 100-per-cent. modulation, in which the carrier amplitude is varied at a very low frequency and gives rise to side bands in exactly the same way. In the case of high-speed signalling by automatic equipment, the sideband spread becomes quite large, demanding adequate attention.

The wave form of a Morse signal will be rectangular. Now, the Fourier analysis shows that such a wave form contains high harmonics up to a theoretically infinite limit, and that if these are not all fully transmitted the received signals will become rounded. Seriously rounded telegraphic signals lose their character and become unsatisfactory for automatic telegraphic recording or printing machines. The absolute limit occurs when the 'dot' wave form approaches a sine curve in shape, which will occur when only the 'fundamental' dot frequency is retained, namely the frequency at which successive dots follow each other. Consider telegraphic signals at 250 words per minute. An average word is said to consist of five letters, each of which averages ten dots in duration, allowing three dots to one dash and a three-dot interval between letters. This signalling speed therefore corresponds approximately to $250 \times 5 \times 10$ dots per minute, or 2,000 dots per second. The 'dot frequency' is thus 2,000 half-cycles per second, or 1,000 cycles per second, and since this corresponds to a minimum sharpness in the received wave form a higher frequency is desirable. Hence it becomes apparent that high-speed telegraphy calls for almost as high a maximum modulation frequency as average speech, and all the precautions necessary to preserve speech quality remain. Damping must not be too low at any point in the transmitting circuits, and too great receiver selectivity must not be used.

Treating telegraphic signals as a form of modulation, the methods used for keying should be considered in this chapter also. Normally, of course, the carrier is completely interrupted by the telegraph key, a process only possible in very low-power transmitters, where the key may be inserted in a lead from the high-tension supply. To key high powers directly, the key, or an incoming line signal from a telegraph circuit, operates a relay.

The larger current thus controlled operates a power relay or quick-acting electrically driven power switch, which may close the main high-tension circuit or even an aerial-feeder. In the earlier high-powered transmitters, where it was thought necessary to key the power circuits themselves, this problem was a difficult one. Costly and sometimes troublesome oil-immersed switchgear became necessary, and the speed of keying was limited by the inertia of the moving parts. These difficulties led to the frequent adoption of a system equivalent to frequency modulation, in which the keying relay merely short-circuited a few turns of a tuning inductance, or otherwise slightly changed the radiated wavelength, leaving the aerial power unchanged. It was then possible to place the relay at a point where the current to be broken was smaller. The signal then consisted of two slightly different waves, known as the 'spacing' wave and the 'marking' wave. A heterodyne receiver would give a different audible beat with these two, and might be tuned to the marking wave in addition, thus enabling the signal to be read as a changing note.

A second system, now almost abandoned except for low-powered or mobile use, was termed 'modulated continuous-wave' or 'interrupted continuous-wave' signalling. Here the steady carrier is either modulated by a suitable low-frequency note, which might be produced by an alternator in series with the high-tension supply, or interrupted at an audible frequency by some form of commutator. The resulting 'tone-modulated' signal is very easy to read by ear, and the key is inserted at some point in the modulating device where power is not too great. The system is useful to-day on the ultra-short waves, where selectivity is often poor, but is no longer allowed in the busy commercial wave bands because of the interference which the side bands of the modulation can cause to adjacent services. It is also in use as a substitute for the older spark transmitters employed on board ship. A carrier having marked alternating-current ripple due to bad smoothing of the high-tension supply may sometimes be heard and approximates to a modulated continuous-wave transmission.

Neither of these systems is ideal for high-speed signalling, and modern practice tends to favour the complete interruption of the carrier between signals. Keying, however, can be carried out in low-powered circuits in the modern stabilized trans-

mitter by the simple process of effecting it in one of the earlier stages. If all the amplifying or frequency-multiplying stages in the transmitter employ class B or C biasing conditions, then in the absence of 'drive' from the master oscillator no anode current will flow, and the whole transmitter will be dead. It would be possible to take advantage of this fact by keying the master or crystal oscillator, where power is low. A key could be inserted in the anode supply, or alternatively grid keying could be used, in which the closing of the key serves to remove from the oscillator grid a high negative bias which maintains the valve inoperative during spacing periods.

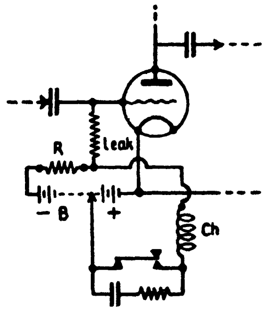


FIG. 100. Grid Keying.

Whilst keying of the master oscillator is sometimes used, it is unwise to do this when avoidable, since the frequency stability of this stage is likely to be impaired by the disturbance. It is better to leave the oscillator untouched and instead to key one of the following amplifying or doubling stages. Here grid keying is very suitable, and this is indicated in Fig. 100. The battery or other source *B* provides a high negative bias, sufficient to cut off all anode current and thus prevent any radio-frequency drive from passing to the following stages and thence to the aerial. On closing the key or keying relay, this bias is removed, leaving only the correct operating bias, and the transmitter at once radiates normally. The resistance *R* is to prevent undue load upon the battery while the key is closed, whilst the choke keeps stray radio-frequency potentials from the key. A condenser in series with a small damping resistance may be joined across the contacts to absorb any sparking, this being more often necessary if the key be used to break heavier currents than in the circuit shown. It is of course equally possible to key the anode supply to the stage, and a number of other methods of little technical interest are used. An excellent modern method used where the keyed stage employs a pentode is that of suppressor-grid keying, the circuit being similar to that shown. All these grid methods are characterized by power economy and are suited to simple keying relays. They are termed low-level or low-power keying, as opposed to the old high-power methods.

Similarly, modulation for telephony can be carried out at either high or low power. Early transmitters modulated a power oscillator directly, and whilst this can yield very fair quality if the frequency stability be inherently good it is now seldom used. Instead, the modulation methods we have outlined are applied either directly to the power output stage of a driven transmitter, which is termed 'high-level' modulation; or alternatively to one of the preceding stages, which is 'low-level' modulation. It is difficult to say which method is the better, although the tendency has been to adopt low-level methods in recent years. Very recent station designs, however, have reverted to high-level, taking advantage either of series modulation or of the newly developed large class B modulator stages. At first sight it would seem that high-level modulation would be wasteful owing to the power consumption of the modulators. Originally, when modulators were class A biased and choke coupled, this was the case; but the superiority of low-level methods has now largely disappeared. It is found that the power used in high-level modulators is no more with efficient design than the extra power which has to be provided for in the final radio-frequency amplifier of low-level systems in order to allow for the modulation peaks which this stage must handle. If these peaks are of the same amplitude in the two cases, it is shown that theoretically the power consumption will be very similar.

One advantage of low-powered modulation over all others is that 100 per cent. modulation is easily obtainable under practical conditions, although it is always a fundamental that distortion increases somewhat as modulation is increased. As a general rule a maximum of between 70 and 90 per cent. is considered sufficient, both to maintain good quality, to reduce risk of momentary over-modulation, and for practical reasons. In the case of anode systems, for example, curvature of the modulator-valve characteristic makes it impossible to reduce the anode current to zero without leaving the linear region. Since this would be essential for 100 per cent. modulation, it places a limit on the depth that can be used. When modulating a low-powered driving stage, however, it is possible to modulate fully that which follows without fully modulating the driving stage. Suppose that the driver can deliver twice the power or potential swing necessary to excite the following amplifier fully, and that this

is biased class C so that when the drive falls to 50 per cent. of its mean value radio-frequency output ceases. Then 50 per cent. modulation of the driver output will change the excitation of the following amplifier from that which gives zero output to double the normal, full modulation being achieved. This can only be done because the modulated stage has a greater maximum output than necessary, only some half of this being required to yield normal drive.

Summing up, it may be said that whereas it is difficult to modulate the final power amplifier of a transmitter fully, we can reach an equivalent effect by lightly modulating one of the driver stages preceding it. The final stage should now be adjusted as a class B amplifier. It now amplifies the modulated signals from preceding stages, and can be fully driven by less than their full output, thus changing a partially modulated signal into a fully modulated transmission.

In conclusion some mention must be made of controlled or quiescent carrier working. This is a system by which power is economized and interference to other channels reduced. The transmitter is so arranged that little or no carrier is radiated when no modulation occurs, but increases to a value proportional to the modulation depth during speech. In this way only the minimum energy is radiated at any time to meet the needs of the moment, low modulation occurring upon a low carrier and high modulation upon an increased carrier. To effect this the carrier must be controlled by the mean speech or modulation level. Many methods are possible. As an example, the amplified speech may be rectified by a diode circuit of time constant some one-tenth second. This produces a bias potential which is proportional to the mean speech amplitude. This bias may now be applied in opposition to that of an over-biased carrier amplifying stage, so that, as the speech amplitude increases, bias is reduced and the output from the stage correspondingly increased. The whole carrier is thus controlled by the mean speech level and can be arranged so that no radiation at all occurs in the absence of modulation.

EXAMINATION QUESTIONS

1. Give a diagram showing how a valve transmitter can be modulated by speech currents for telephone transmission, and explain briefly the action.

City and Guilds of London Institute. Grade 1 1929.

2. What exactly is meant by the statement that a transmitter radiates 5 Kw. unmodulated ? What is the maximum power radiated in this case ? How is each of these powers affected by modulation ? Give reasons for your answers.

Institute of Wireless Technology. November 1935.

3. Describe, giving diagrams of connexions, methods of applying modulation to the high-frequency carrier of a transmitter.

I. W. T. November 1934.

4. Sketch the modulation envelope of a carrier wave under the following conditions:

(*a*) 100 per cent. modulation by a single sinusoidal frequency.
(*b*) About 130 per cent. over-modulation by the same frequency.
(*c*) About 50 per cent. modulation by a rectangular wave form. In this case, what would be the sideband frequencies produced ?

5. What is the meaning of the following terms:

(*a*) Frequency modulation;
(*b*) Single sideband amplitude modulation;
(*c*) Grid-bias modulation;

and what are the advantages and disadvantages of each ? What is the effect of modulating a wireless transmission by a frequency higher than the carrier frequency ?

6. Give a circuit diagram and description of a simple radio-telephone transmitter having a crystal-controlled drive. Describe in detail the method used to effect modulation. *A. I. W. T., June 1937.*

7. Explain the action of a simple carbon microphone, and give a circuit for its connexion to the input of a valve amplifier. On average speech a microphone produces an average peak potential of 0·1 volt across the primary of a step-up transformer of ratio 100 : 1. This feeds an amplifier which in turn feeds the grid of a modulating valve having a negative grid bias of 200 volts. If this valve works as a class A or linear amplifier, what must be the over-all voltage amplification of the modulation amplifier used, measured from the grid of the first valve to the anode of the last ?

8. What is meant by (*a*) choke control, (*b*) series modulation, and (*c*) class B modulation ? What are the principal advantages of each system ?

9. What is the saving in power if a telephony transmitter radiates (*a*) the carrier wave and only one sideband, (*b*) one sideband alone? What steps must be taken at the receiver to receive each of these transmissions without undue distortion?

10. State what you know of the methods adopted to provide secrecy in radio-telephone conversations.

11. Describe three methods of keying a radio-telegraphic transmitter.

12. What is meant by (*a*) high-level modulation, (*b*) low-level modulation? Mention the chief advantages of either system.

13. What do you understand by 'the sideband theory'?

14. What are 'side bands' in radio-telephony? Discuss their existence as physical realities at the various stages of a radio-communication channel. See also Chapter XI.

CHAPTER X

RECEIVING CIRCUITS

We must now consider more fully the circuits by which the signals are to be conveyed to the detector and the telephones. We will take firstly the simple early form of receiver in which energy from the aerial is rectified without previous amplification. In considering any wireless circuit it is necessary always to keep in mind that we are dealing with high-frequency currents and that stray capacities present a comparatively small reactance to such currents, and may therefore cause troublesome leakage paths.

Thus, stray capacities must be avoided at points in the circuits where there is a comparatively high potential. For example, in the simple circuit of Fig. 66 it is important that the potential developed in coil L should be applied direct to the crystal and the telephones. If telephones were inserted before the crystal, the capacity to earth of their windings would provide a partial short circuit through which signal currents would leak to earth without passing through the crystal.

These considerations are all the more important in view of the fact that the high-frequency currents and voltages in the received signals are very minute. Indeed, the energy available for working the receiving apparatus usually lies between one-hundredth and a millionth of a millionth of a watt (i.e. 10^{-8} to 10^{-12} watts). It is obvious, therefore, that in designing any receiving apparatus every possible channel by which the small energy available may be lost must be eliminated. Thus, in all receiving circuits, in addition to avoiding stray-capacity effects between leads, &c., insulation must be very good and every care must be taken to reduce the resistance losses of coils and condensers to a minimum.

Assuming that an elevated aerial with ground connexion is used, we have seen that the aerial may be regarded as a condenser, and that the earth connexion and the inductance coil will introduce into the circuit a certain ohmic resistance. The aerial circuit of the simple receiver shown in Fig. 66 may be represented by a circuit containing a condenser, a coil, and a resistance in series (see Fig. 101). The self-induction of the

aerial wire is assumed to be included in the coil L, and all the various ohmic resistances of the aerial and coil and earth connexions and the insulation losses included in R. The high-frequency oscillating potential set up by the signal between A and B will cause a current to flow in the circuit. Let us suppose the maximum value of the potential to be V and that of the current to be I. Then, from what we have seen in Chapter II,

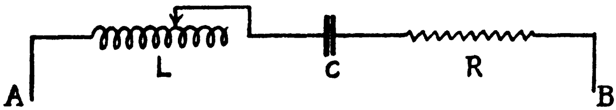


Fig. 101.

this applied voltage has to overcome a back E.M.F. due to the self-induction of the coil equal to ωLI, a voltage of $I/\omega C$ across the condenser, and an ohmic pressure drop IR across the resistance R. We have also seen that in a coil and a condenser the current respectively lags behind and leads the applied voltage by 90°, and that in order to obtain the sum of these three voltages, which are balanced by the applied voltage, they must be added together geometrically and not arithmetically.

Hence the maximum value of the current developed is connected with that of the induced voltage by the relation (see Chapter II, page 32)

$$I = \frac{V}{\sqrt{\left\{R^2 + \left(L\omega - \frac{1}{C\omega}\right)^2\right\}}}. \tag{1}$$

In order that the current developed for any applied potential of a frequency $f = \omega/2\pi$ may be as great as possible, the capacity C and the inductance L must be adjusted so that

$$L\omega = \frac{1}{C\omega} \quad \text{or} \quad \omega^2 = \frac{1}{LC}.$$

From the above expression it is clear that the smaller the value of R the greater will be the current for a given resonant frequency ω. Hence the importance of keeping down losses is evident.

But we shall see that the effect of R is still more important when we take into consideration the fact that the signal we may wish to receive is not the only one travelling in the ether, and that the aerial is likely to be affected by other waves of

different frequencies. Supposing another wave, of frequency ω_1, affects the aerial: then, if V_1 and I_1 are the maximum values of the voltage and current due to this wave,

$$I_1 \quad \frac{V_1}{\sqrt{\left\{R^2+\left(L\omega_1-\frac{1}{C\omega_1}\right)^2\right\}}}.$$

In this case $L\omega_1$ will not equal $1/C\omega_1$, and therefore, if the interfering transmitting station is so placed that the voltages applied by the two waves are equal (i.e. if $V = V_1$), I_1 will be less than I. If, however, ω_1 is not very different from ω, it is clear that if R is not small the decrease in the aerial current will not be very great. Hence the detector and telephones will be strongly influenced by interfering waves. In other words, the receiver will not be selective, a subject to be discussed in the next chapter.

Now the number of wavelengths available and suitable for various radio-communications is limited, and therefore selectivity is a most important element to be taken into account in the designing of receiving apparatus. We have seen that in the simple aerial circuit dealt with this is attained by reducing the ohmic resistance. But in such a circuit this resistance cannot be eliminated altogether. In order, therefore, to make further use of the effect of resonance, which depends largely on the resistance losses being small, the detector is not directly connected across the coil L in the aerial, but may be connected across a circuit composed of a coil L_1 and condenser C_1 which is coupled to the coil L in the aerial circuit. This arrangement is shown in Fig. 102. In particular, great care must be taken to design the coils so that the self-capacity effect between the various turns and the losses in the insulation of the windings and supports of the coil are as small as possible. Otherwise the capacity effect will provide a path for the high-frequency currents, and the potential built up across the coil will be reduced. The oscillations across the coil L in the aerial circuit are transferred to the circuit $L_1 C_1$ by the mutual induction of the coils L and L_1 or by one of the other forms of coupling dealt with in Chapter IV. The current oscillations in the circuit $L_1 C_1$ build up a high-frequency alternating potential across the condenser C_1 which is applied to the crystal. The most convenient adjustment for coupling between L and L_1 must be found by experiment. An effect of extremely tight coupling is that the secondary circuit may apply a large

damping effect to the aerial circuit and may draw so much energy away from it that the energy available in the received waves may be insufficient to set the aerial in free oscillation.

When still more selectivity is desired, an intermediate circuit, consisting of a coil coupled to the aerial, a condenser, and another coil coupled to the circuit $L_1 C_1$, may be employed.

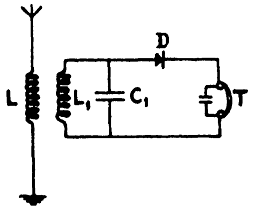


FIG. 102.

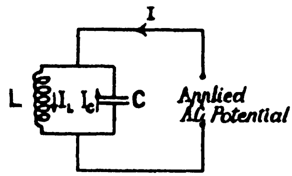


FIG. 103.

Even with such an arrangement, however, interference from near-by stations may still be very troublesome, and another arrangement called a 'rejector circuit', or sometimes a 'wave trap', may be employed. The action of this type of circuit is somewhat different from the circuits already described. Let us consider an oscillating potential of maximum value V and frequency f ($= \omega/2\pi$) applied externally across a circuit consisting of a coil of inductance L and a condenser of capacity C (see Fig. 103). Assuming that the circuit LC has no resistance, there will then be an oscillatory current in the coil, whose maximum value will be given by $I_L = V/\omega L$, *lagging* 90° behind the applied potential.

Similarly, there will be an oscillating current through the condenser given by $I_C = \omega C V$ which will be *leading* the applied voltage by 90°.

I_L and I_C are thus 180° out of phase, and the external current will therefore be given by

$$I = I_L - I_C$$

$$= \frac{V}{\omega L} - \omega C V$$

$$= V\left[\frac{1}{\omega L} - \omega C\right]. \tag{2}$$

If the circuit LC is in resonance with the frequency of the applied potential, i.e. tuned to the frequency $f = \omega/2\pi$, we have seen

that $\omega^2 LC = 1$. For this case the external current I given by (2) will be zero, and the circuit LC acts as if it had an infinite impedance or resistance to the current of the particular frequency with which it is in resonance.

The application of the applied E.M.F. across the circuit, however, builds up a large oscillating current in the circuit itself, whose value is given by

$$I_L = I_C = \frac{V}{\omega L} = V\sqrt{\frac{C}{L}}.$$

If the circuit is not adjusted to be in resonance with the frequency of the applied oscillation, then I_L is no longer equal to I_C, and a current will flow through the circuit and the external circuit. If the frequency for which the circuit is tuned is greater than that of the applied potential, the current passing through the circuit will flow mainly through the coil, and if the frequency of the circuit is less than the applied potential the current will flow through the condenser. Also the greater the capacity of the condenser, and the smaller the inductance of the coil, the greater will be both ωCV and $V/\omega L$, and in general the greater will be their difference. Hence the *greater* the ratio C/L the *less* is the impedance of the circuit to oscillating currents at non-resonant frequencies.

If the circuit possesses some resistance R, it can be shown that the effective impedance to an applied potential at the resonant frequency is given by L/RC ohms. This is a most important quantity, and is termed the *dynamic resistance* of the circuit.

Since the leading and lagging current components have been shown to be equal at resonance, the external current being zero, there will be no change in phase when an alternating potential is applied across a resonant circuit. The circuit thus behaves exactly as a pure resistance of value L/CR ohms, namely of the dynamic resistance. It is sometimes important to realize that a change of phase will occur directly the circuit is detuned from the applied frequency, and this change will be opposite on either side of resonance. Thus if the circuit be tuned to a higher frequency than the applied potential a greater current will flow through the inductive branch, and therefore the current will on the whole be lagging; and vice versa. A reversal of phase will occur if we tune a circuit through resonance from above to below the frequency of an applied potential.

Dynamic resistance also determines selectivity. As the ohmic resistance of a circuit is reduced, the expression L/CR will increase, whilst its impedance at non-resonant frequencies decreases. The ratio of response to the tuned signals relative to other signals on adjacent frequencies thus becomes better. As an illustration, an efficient circuit may have an ohmic resistance of 1 or 2 ohms only, whilst the dynamic resistance may exceed 150,000 ohms.

It so happens that at all but the very highest frequencies most of the losses and resistance which make up the term R will be present in the coil. The condenser by which it is tuned will, if it be a good one having an air dielectric, contribute very little to the total circuit damping. Let us neglect this altogether. Then by substituting the expression $1/\omega^2 L$ for C, as we may do at resonance, the expression L/RC for resonant impedance becomes equal to $\omega^2 L^2/R$. Now both L and R are properties of the coil in use, and it will be noticed that the ratio of L to R determines the impedance. A coil of high L and low R will have a high impedance at resonance, and a low impedance at other frequencies. It thus results in a 'good' tuned circuit, of high natural selectivity; whereas a low ratio of L to R will imply a circuit of poor properties.

It has been found convenient to assign a symbol to this ratio, and we have already noted that the letter Q has been given to the expression $\omega L/R$. Thus since $Q = \omega L/R$, the formula governing resonant impedance of a circuit formed of a coil of 'goodness factor' Q tuned by a perfect condenser would be

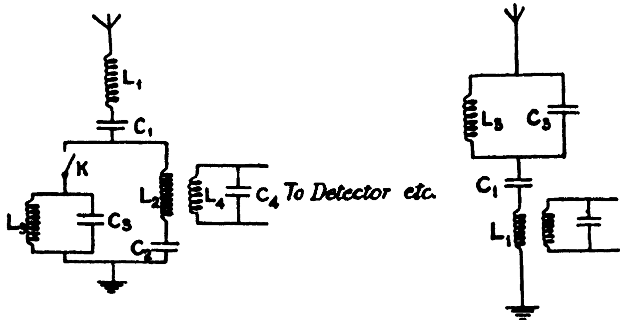
$$\text{dynamic resistance} = L/RC = \omega^2 L^2/R = \omega LQ.$$

In practical radio-engineering work this factor Q is found a very convenient term by which to describe the general quality of coils or tuned circuits. Whilst not a fundamental quantity like L or C, it provides just that convenient term that is needed in practical circuit design, since a 'high-Q' circuit will be selective and will offer a high dynamic resistance across which large oscillatory potentials can be built up, whilst a 'low-Q' circuit will have just the opposite properties. In American papers Q is freely used, and it will be useful to form an idea of typical values that may be encountered.

At broadcast or medium frequencies, of some 1,000 Kc., a Q of 100 or less implies an average or poor circuit. Figures of from

150 to 200 are considered good, whilst a Q of 300 can only be attained by specially efficient methods of construction. The quartz-crystal resonator, which is one of the most efficient devices known, may have a Q of 30,000. It is therefore very lightly damped and highly selective.

We must now return to study the methods by which an



(a) FIG. 104. (b)

efficient circuit can be used to reduce interference from any particular signals, in the form termed a rejector circuit or wave trap. This arrangement was at one time common in receivers, but the latter are now sufficiently selective without it, and wave traps are confined to special uses, such as cutting out interference from a local transmitter.

One method of using a rejector circuit to reduce interference is shown in Fig. 104 (*a*). The rejector circuit L_3C_3 is tuned to the frequency of the signal to be received and can be inserted by closing the switch K. The portion of the aerial circuit L_1C_1 is tuned to the same frequency. Similarly, with K open the circuit L_2C_2, which is known as an acceptor circuit, is tuned to the same frequency.

When these two circuits L_1C_1 and L_2C_2 are connected together in series, the whole circuit $L_1C_1L_2C_2$ will be tuned for a frequency given by

$$f = 1\Big/2\pi\sqrt{\left\{(L_1+L_2)\left(\frac{C_1C_2}{C_1+C_2}\right)\right\}},$$

$$f = 1/2\pi\sqrt{(L_1C_1)}, \text{ since } L_2C_2 = L_1C_1.$$

Hence the tuning of the aerial is not altered by connecting the two circuits in series.

Suppose, now, K is closed: then any oscillation of the fre-

quency to which the aerial is tuned will find an easy way to earth through the acceptor circuit L_2C_2 and in doing so will induce oscillations in L_4C_4, while the rejector circuit, assuming that its ohmic resistance is low, will present a very high impedance to such signals. On the other hand, any interfering signals of different frequency will find an easy path to earth through the rejector circuit, while at the same time the acceptor circuit L_2C_2 presents a high impedance to such signals. Thus the interfering oscillations are prevented to a large extent from reaching the receiver circuit L_4C_4.

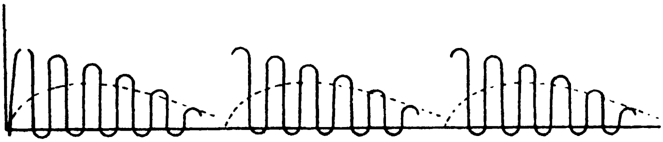
When a particular signal only is to be eliminated, a rejector circuit such as L_3C_3 can be introduced in series in the aerial circuit as in Fig. 104 (*b*). In this case the rejector circuit is tuned to the frequency which it is desired to cut out. The circuit then imposes a high impedance for waves of this particular undesired frequency, which are accordingly rejected.

For carrying out the adjustment and tuning of the aerial and receiving circuits, oscillations are induced in the circuits from another circuit consisting of a coil of known self-induction and a variable condenser of known capacity. By the variation of the value of this condenser, and the calibration chart provided, the circuit can be adjusted to be in resonance with any desired frequency or wavelength. The arrangement is known as a wavemeter. If the coil is placed near a transmitting circuit, and if some form of detector with a galvanometer or a pair of telephones is connected across the condenser, the wavelength emitted can be measured by adjusting the variable condenser until the sound in the telephones or the deflexion of the galvanometer is a maximum. The wavelength can then be deduced from the formula $\lambda = 1885\sqrt{(LC)}$ or from the calibration chart of the instrument.

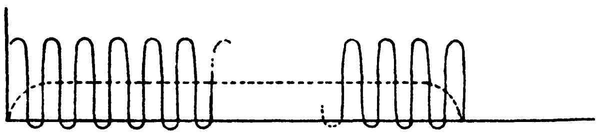
The converse process is used for the adjustment of the receiving apparatus, the circuit of the wavemeter being energized so as to emit the frequency to which it is desired to adjust the receiver. A convenient way of doing this is by means of a 'buzzer', or a valve oscillator may be used.

Hitherto we have dealt with the reception of damped waves, in which a dash or dot is made up of a series of oscillations of decreasing amplitude, each group of which charges the telephone condenser. In the case of continuous waves, however, a dot or dash does not consist of a series of trains of waves, but of an

unbroken series of high-frequency oscillations extending over the period during which the sending key is pressed. The rectified current corresponding to a signal due to a damped wave is shown in Fig. 105 (*a*), and that due to an undamped or continuous wave in Fig. 105 (*b*). The dotted lines represent the change of current through the telephones as already explained. If



(*a*)



(*b*)

FIG. 105.

we examine Fig. 105 (*b*), we see that the current through the telephone corresponding to the signal is steady, except for an increase at the beginning and a decrease at the end. Hence the only response the telephone diaphragm can give is a click at the beginning and end of the signal. We see, therefore, that rectification by a crystal only will not allow continuous waves to be received. The earlier methods of receiving such waves consisted of the breaking-up of the waves into groups, by either mechanical or electrical means. These methods were never entirely satisfactory, and have given place to the 'heterodyne' method of reception.

In principle, the method depends on the production of 'beats' between two waves of slightly different frequencies. The receiving circuit is excited by a small 'transmitter' coupled and adjusted to produce oscillations differing very slightly in frequency from those of the incoming signal. Since the local transmitter is producing continuous waves, its oscillations will not be perceptible in the telephone. Waves from the distant transmitting station, however, will combine with the local oscillations, and two sets of oscillations will be produced in the receiving circuit: (1) a steady continuous oscillation whose amplitude is equal to the difference between the amplitude of the signal and the locally

produced oscillation; (2) an oscillation having an amplitude varying between zero and twice that of the signal. The former will not be audible, but the other will have a frequency f_1-f_2, where f_1 is the frequency of the signal and f_2 that of the local oscillation, and if f_1 and f_2 are nearly equal this latter will be sufficiently slow to influence the diaphragms of the telephones. Thus, if we wish to receive continuous-wave signals having a frequency of 100,000 cycles per second (i.e. signals corresponding to a wavelength of 3,000 metres), the local transmitter may be tuned to produce either 101,000 cycles per second or 99,000 cycles per second. The telephone will then respond to the beat oscillation of frequency 1,000, resulting from the interference of the two former frequencies in the receiving circuits. This effect is analogous to the beat heard by the ear when two similar musical tones are played together.

As the local oscillator is tuned to different frequencies, the low-frequency beat heard in the telephones will vary, and the receiver can therefore be adjusted so that the audible frequency can have any desired pitch. If the local oscillator is exactly tuned to the same frequency as that of the incoming wave, then no beat note will be produced and there will be silence in the telephones. This property is employed in the heterodyne wavemeter for the adjustment of circuits. At zero beat, and if the two oscillations are in phase, the local oscillator will reinforce an incoming carrier. This condition has been termed 'homodyne' reception.

It has been stated that the output current of most detectors is proportional to the square of the amplitude of the received current. With heterodyne reception the loudness of the received signals can be shown to be proportional to the product of the amplitude of the received signal current and the current due to the local oscillator. The strength of the signals in the telephones is therefore proportional to the strength of the received current instead of to its square. Heterodyne reception is accordingly much more sensitive than simple detection for weak signals. This fact is most important.

The fact that the note of the received signals can be adjusted by the operator to any pitch gives the method a great advantage over other methods of reception; for the note of the received signals can usually be so adjusted as to be different in pitch from the signals from other stations or from atmospherics.

So far it has been assumed that the incoming signals are rectified directly by a crystal or diode detector, because this treatment brings the fundamental aspects of reception clearly into view. To-day, however, this is seldom done, except for very local broadcast reception, or for the measurement of field strength close to a transmitting aerial. Most receivers employ valve amplification, the principles of which have been explained in Chapter VIII; if amplification is omitted a triode or multi-electrode detector valve is used, rectifying as has been described in Chapter VII. Valve detection introduces several advantages, apart from mere consistency. The signal is applied to the valve grid, where it exerts a controlling influence on the anode current. The energy to actuate the telephones now comes from the anode current, and not directly from the aerial circuit. Less load can therefore be placed across the aerial coil, damping is reduced, and selectivity is improved. The circuits previously described have historical interest and illustrate important basic principles, but they have been largely replaced to-day by those we shall now consider.

The simplest modern receiver thus becomes a one-valve detector arrangement, which may be followed by a low-frequency amplifier whenever necessary to make signals louder or to drive a loud speaker, without modifying the action of the detector in any way. The principle of low losses, which has been discussed in relation to the crystal, still applies in this and all other circuits, but to a somewhat lesser extent. Losses can be partly overcome by the use of reaction, the principles of which were given in Chapter VI. The Hartley circuit of Fig. 58 (*a*) makes an excellent receiver for short waves, whilst on longer waves the original type of circuit used was magnetically coupled, as in Fig. 52. This is the simplest and most fundamental triode-valve receiving circuit.

Reaction is the one factor which makes the simple valve detector useful even after so many years of progressive development. We have seen that by reaction a certain amount of energy can be fed back from the anode to the grid circuit, thereby reducing the losses or damping of the latter from all causes. The degree of this reaction coupling is adjusted in reception by means of the variable magnetic coupling of Fig. 52 or the variable condenser C_2 in the Hartley circuit. As reaction coupling is slowly increased from zero, the grid-circuit losses

are reduced, and hence the potentials built up across the tuned circuit LC increase. Since this is coupled to or forms part of the aerial circuit, losses in the aerial are also partly compensated. We should, however, bear in mind that reaction cannot actually remove such losses but can reduce their effect of weakening the signal or flattening the resonance curve. It is always best to maintain losses at a minimum in the first place, thus not expecting too much from the reaction effect, which is limited in its scope by eventual instability.

A little reaction coupling therefore increases signal strength and selectivity, at the expense of energy drawn from the anode supply source. We can further increase reaction with further improvement in these two factors, until a state is reached when the energy replaced by its use nearly equals that dissipated by the combined effects of all circuit resistance and losses, represented by the collective factor R in the preceding equations. Any further increase in coupling will now make R negative, which has been shown to be the condition for continuous oscillation. The receiver will therefore oscillate at approximately the resonant frequency of the grid circuit, although at first these oscillations will be of small amplitude. Increased reaction would increase this, making the circuit virtually a small transmitter. However, in the reception of telephony it is not desirable that the receiver should oscillate, since distortion results.

With reaction set critically just short of the oscillation point, the simple receiver becomes very sensitive. Its selectivity is adequate for many uses when interference from adjacent stations is not too great, and we shall consider its actual value later. Thus the receiver is suited to a large class of all-round reception, which together with its cheapness and simplicity accounts for a wide degree of popularity. The circuit becomes increasingly useful at the shorter wavelengths, where more complex types are difficult to design and the wide frequency separation between most stations makes high selectivity unnecessary. It forms a major class of receiver amongst amateurs, the public, and the less-exacting types of commercial services. Without reaction, however, it would be little better than a good crystal receiver.

The type of detection usually employed in a receiver of this kind is of the grid-leak and condenser variety, on account of the sensitivity of this circuit to weak signals. Since no amplification

is used before the detector, signals will often be relatively weak, and it is most important that they be rectified as perfectly as possible. We have already discussed the most important systems of detection. The selection of any one of these is determined by factors in the remainder of the receiver which will be pointed out as they arise, or by reasons of quality, which will be discussed in the following chapter on reproduction. It will not therefore be necessary to devote special attention to detection in the present chapter.

The valve detector will, of course, deliver an audible signal from any incoming wave which is modulated by telephony, or in any other way, or from a damped wave of the types now nearly obsolete. We have seen that it will not do so from a continuous wave, but that heterodyne by a separate oscillator will make this audible. The reactive detector overcomes this defect also, since if reaction be increased until the valve oscillates gently, this oscillation itself can beat with the incoming wave to produce an audible note. It is necessary to detune the receiver slightly, to such an extent that the oscillations produced will differ from the incoming signal by the desired beat frequency, perhaps 500 to 5,000 cycles per second. At short wavelengths this detuning represents such a minute fraction of the high incoming frequency that it has very little effect in weakening the potentials built up by signals across the grid-circuit reactance. Reception is thus very efficient, and since the local oscillation adds energy to that of the signals the arrangement shows excellent sensitivity. The background noise and other circuit noises are naturally a minimum in the case of such a simple single-valve circuit, and it is claimed by many that this circuit is the most sensitive arrangement possible for the reception of the weakest continuous-wave signals by ear. It is inferior to many others, however, for commercial purposes, which normally demand a stronger output than that just audible to the trained ear.

On long wavelengths, where the signal frequency is much lower, the detuning necessary to produce an audible note becomes a larger proportion of the whole. Loss of signal strength will then occur, becoming so great towards the longer wavelengths that self-heterodyne, or autodyne, as it is also termed, is no longer satisfactory. The separate heterodyne is used in this case, and local oscillations are induced into the receiver by a valve which is acting as a small generator. It is more convenient

for this purpose to maintain the locally generated oscillations in a tuned anode circuit rather than in the grid circuit, as described above. The principle of the method, however, is similar. The local generator may be coupled to the receiving circuits electromagnetically, as shown in Fig. 106, or may merely be in proximity to the receiver.

The local oscillations must not be so strong as to paralyse the

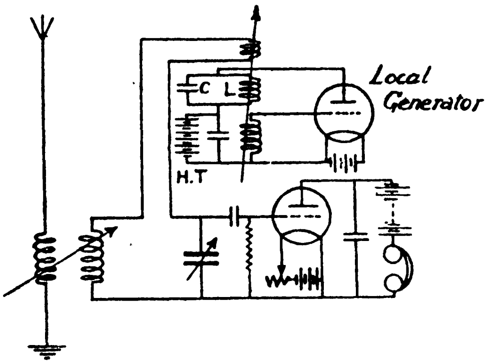


FIG. 106. Separate Heterodyne.

valve: hence in heterodyne oscillators the anode voltage is usually much less than in receiving circuits. In fact, with some valves it is sufficient to arrange that the potential of the plate is above that of the filament and grid by the amount of the voltage of the filament battery only, and no additional high-tension battery need be provided.

We have seen that the majority of detectors operate according to a square law, particularly when the signals are weak. Even the diode would be unsatisfactory in the case of a very low signal level, because in all valves a certain amount of noise is generated, and this may mar or even swamp the weaker signals. One source of this noise is termed Schott effect, being due to irregular emission of electrons from the heated cathode. A second is the Thompson effect, which is not confined to valves, but occurs in all circuits, particularly those in which resistance is high. This is a fundamental effect due to the motion of the electrons which constitute an electric current through a resistance. Current is regarded as a uniform flow of a very large number of electrons, but in actual fact these electrons do not move with exactly equal velocity at all times. If a given current

flows through a conductor we may say that *on the average* a number of electrons n will pass across a boundary or imaginary section at one end of the circuit whilst within the same period n electrons must also pass across the corresponding section at the other end. This follows because current flows through the circuit, and there is no accumulation of electrons within.

If the current be weak, the number of electrons will be smaller. Also, if we take a very brief time interval the number passing over any section will be less. In the case of a weak current for a brief time, therefore, the number of electrons may be too small for their varying velocities to exhibit a precisely constant average, and the number of electrons passing in and out of the circuit will differ slightly from one brief time period to the next. This means that the current is in fact slightly irregular, a necessary corollary from the fact that it is composed of a large but not infinite number of discrete particles, the electrons. A slightly irregular current flowing through a high resistance will give rise to irregularities of potential across its ends. This occurs in a grid leak, for example, and causes a slight fluctuating potential upon the grid, noticeable as a hissing sound after sufficient amplification.

This Thompson effect is fundamental, and is the limiting factor in all electrical amplifying circuits. We can amplify until the Thompson noise inherent in the first grid circuit becomes appreciable, after which no further amplification will be of any help in raising a weak signal to audibility. The noise potential at a typical detector grid may be of the order of 5 microvolts; it is often more, and may be added to by other noises due to faulty components or electrical interference from outside. Hence unless the signal exceeds this value it will be marred by noise, and no amount of amplification following the detector can overcome the difficulty. For this reason the sensitivity of a receiver can only be increased beyond a certain point if amplification be used before it, namely, at the frequency of the incoming signal.

These facts give the clue to the relative use of low-frequency and radio-frequency amplification in the modern receiver. Low-frequency amplification cannot improve reception except to make weak signals stronger, but it necessarily increases all noise and interference present at the detector to an equal extent. At one time, when radio-frequency stages were very inefficient,

three or more low-frequency stages were sometimes used after the detector, particularly if this was a crystal. But to-day this is seldom done. Signals are amplified before detection, so that even from the weakest station a fairly high level exists after detection. The method is in fact often termed 'high-level detection' to distinguish it from the older systems of 'low-level detection'. A minimum of low-frequency amplification is then necessary, thus keeping noise at a minimum and avoiding distortion. For telephone listening a single stage will be ample, and telephones are often used directly after the detector. When a loud speaker is employed, two amplifying stages are usual, one only being sometimes found sufficient. It is only in special circumstances, such as the commercial operation of automatic recording machines, that a larger number of stages may be needed. Here it may be necessary to make up for amplitude lost in selective filters, secrecy devices, or telegraph cables. The recording devices may need a very high signal amplitude for proper operation, whilst noise may be less important than when the ear is used. Thompson noise will also occur in the grid circuit of signal frequency amplifiers, but since the resistance of the coils used is low, the effect is less serious.

The principles of low-frequency amplification have been reviewed already. By far the most important remaining consideration in its use is that of quality in reproduction, the design of stages where this is unimportant being very simple. This problem is so bound up with the general question of quality, which forms the subject of the next chapter, that it will be best left for consideration therein. We will now study the more important problem of radio-frequency amplification, with particular reference to short waves; but before doing so there remains a special detector circuit of exceptional interest which should be discussed. This is known as super-regeneration, and is a method of detection evolved by Armstrong in America about 1922. It aims at making fuller use of reaction than is usually possible, depending upon the following facts for its operation. When external oscillations are impressed upon a circuit having a small positive resistance, oscillations are built up in the circuit until a certain definite amplitude is obtained. These oscillations die away when the impressed oscillations are withdrawn. If an impressed voltage is applied to a circuit whose effective resistance is zero, then the circuit will keep on oscillating at its original

amplitude. In the case of a circuit having a negative resistance, the amplitude of the oscillations in it will continue to be built up even after the impressed voltage has been withdrawn. No matter how small the initial impressed voltage may be, the amplitude of the oscillations will continue to grow. At the slightest external shock, a circuit containing negative resistance may become a generator of powerful oscillations which will paralyse the action of the valve. As a rule, therefore, the re-

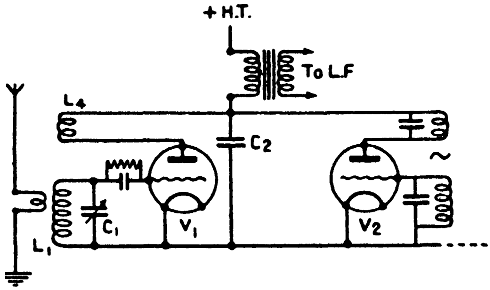


FIG. 107.

active coupling in a receiver must not be so great as to produce a negative resistance, except when receiving continuous waves.

In the Armstrong circuit, however, the grid circuit is given alternately a negative resistance and a positive resistance. During the period of negative resistance a very large current is built up, but just when the circuit is about to begin oscillating on its own account the resistance becomes positive and the oscillations are damped down. Now it is clear that the effective resistance of the grid circuit may be changed either by periodically introducing a resistance into the circuit itself or by varying the effect of reaction by increasing and reducing the anode current: this may be done by periodical changes of the grid or anode voltage of the valve. Both these methods have been described by Armstrong.

Fig. 107 shows how the amount of reaction may be varied by periodic changes in the anode voltage. The coupling between the reaction coil L_4 and the tuned grid circuit $L_1 C_1$ of the detector valve V_1 is increased until the grid circuit is given a negative resistance. Incoming signals are then amplified in the valve V_1 and fed back to the grid by the coil L_4. The valve V_2 is arranged to act as a high-frequency generator at any desired frequency, the circuits being those usually employed for generating oscilla-

tions. The frequency generated by V_2 may be about 10,000 cycles per second. Now the anode circuit of valve V_1 is choke coupled to the anode circuit V_2 by the primary impedance of the low-frequency transformer, through which the rectified signals pass. Hence the anode voltage V_1 is varied at a rate of 10,000 cycles through the application of the voltage generated by V_2. Thus, when the oscillations impressed on the anode of V_1 by V_2 are such as to make the anode voltage of V_1 more positive, reaction is increased beyond the oscillation point and the signal amplitude builds up. Before, however, the valve V_1 can reach continuous oscillation, the voltage applied to the anode by V_2 will reverse and the anode voltage of V_1 will fall, so that the regeneration is checked and, the resistance of the grid circuit being increased, the tendency to oscillation will be removed. The small condenser C_2 provides a moderately low reactance return path to cathode of V_1, without which oscillation would be difficult. V_2 is termed a 'quenching oscillator', since it has the effect of quenching the natural oscillation of V_1. It will be noticed that the quenching oscillations exist across the low-frequency transformer, and will therefore be heard at the telephones or loud speaker. To prevent this a quenching frequency may be selected which is above audibility, say higher than 20,000 cycles per second. If a lower and audible frequency be used to provide more efficient super-regeneration, then a filter may be necessary to keep it out of the telephones, and such filters were usual in early types of receivers. Fortunately only a single frequency is to be eliminated, and so a tuned rejector circuit may be inserted in the low-frequency amplifier for this purpose.

An arrangement for periodically introducing a resistance into the grid circuit itself is shown in Fig. 108. The incoming oscillations are impressed on the grid of the first valve V_1. The second valve V_2 acts as a generator of oscillations at a frequency of about 10,000 cycles, which are impressed on the grid of valve V_1 by the adjustable connexion to the coil L_1. The regeneration by the coupling coil L_4 is carried beyond the oscillation point of the valve V_1. During a half-cycle, when the voltage impressed on the grid of V_1 by the oscillations from the valve V_2 gives the grid an increasing negative potential, the apparent resistance of the tuned grid circuit is negative. During the other half-cycle, however, the voltage impressed from V_2 will give the grid of V_1

a positive potential, and the grid will therefore attract negative electrons to itself. A current will thus flow between the filament and grid, and the part of the valve V_1 between the filament and the grid will now have a resistance of perhaps several thousand ohms. The result of the insertion of a resistance of this value across the grid circuit of V_1 will be to damp out the oscillations

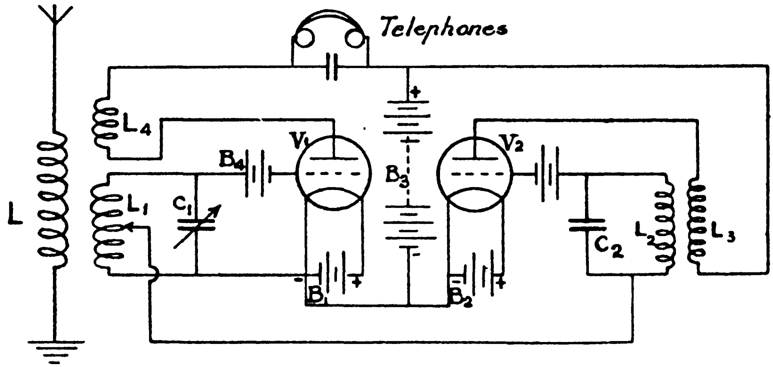


Fig. 108.

in the grid circuit and prevent the valve from becoming a generator. The valve V_1 acts as a detecting valve using anode rectification, adjustment to the correct part of the anode-current curve being brought about by the insertion of a battery B_4 in the connexion to the grid, or by a grid condenser and leak.

The advantages of super-regeneration lie in the fact that during the regenerative periods the signal potentials will be built up to a very large amplitude, which may approach anode-current saturation for the valve used. This will apply equally to very strong or to comparatively weak signals, provided they are not comparable to the noise level inherent in the valve itself or picked up from outside sources; and hence the circuit tends to amplify all signals up to the same level regardless of their initial strength. It has thus a marked ability to self-adjustment of amplification, termed to-day automatic volume control. The enormous amplification produced by the regeneration is not lost during the damping periods, although of course these reduce it to some extent, and the final result is a receiver of perhaps two valves having a sensitivity equivalent to that of many multi-valve receivers of vastly greater complexity.

Armstrong fully realized the extreme sensitivity of his inven-

tion, claiming an amplification as high as ten-thousand-fold from the super-regenerative effect; and recent measurements show that this may not be an exaggeration. He proposed the circuit for broadcast reception, then in a comparatively crude state, and successfully demonstrated reception of distant broadcast stations when using only one or two valves and a small frame aerial. At the time this was a revolutionary performance, and it still remains very noteworthy. Unfortunately, however, it gradually became obvious that with the technique of that period, on medium broadcast or long wavelengths, the circuit had little to be said for it in any respect other than sensitivity. The selectivity was found to be poor, owing to the flatness of resonance naturally associated with the damping periods. This allows interfering stations to break through, and a small frame aerial was necessary to offset this defect. The quality of reproduction was also bad, owing mainly to heterodyne effects between musical frequencies and that of the 'quenching' oscillation. For these reasons the popularity of the circuit declined as the standard of broadcast reception progressed, and little was heard of it for some years, except occasionally in the hands of amateurs experimenting on shorter wavelengths, where high selectivity was not then necessary.

Thus super-regeneration might have remained an historic relic of exceptional theoretical interest, had not the development of ultra-short waves in recent years suddenly raised it to prominence. Certain facts not known at the time of Armstrong's original work were gradually brought to light, partly as the result of amateur effort. One of these was that the amplification given by super-regeneration increased in rough proportion as the wavelength was decreased, until at wavelengths below 10 metres it was comparable to that of the most elaborate receivers. Secondly, it was shown that a relationship existed between the frequency of the incoming signals and that of the best quenching frequency, the two factors increasing approximately in proportion. This meant that whereas on medium wavelengths the best quenching frequency was quite low, well within the musical scale, at short wavelengths the quenching frequency could be raised to perhaps 100,000 cycles, with improved efficiency. At the same time most of the troubles originally experienced from heterodyning against modulation tones, or in filtering out the audible whistle evident in the

telephones of early super-regenerative receivers, were removed. Quality of reproduction then became comparatively good and could be made almost perfect by careful attention to adjustment. Attention to the wave form and amplitude of the quenching oscillation further assisted, and it was found possible to obtain somewhat better selectivity, which whilst still low was adequate for many uses. The removal in this manner of the chief defects of the old receivers, combined with the simplicity, high sensitivity, and automatic volume control action, has made the circuit a very useful one indeed in the ultra-short-wave region; although it still remains of little use above 20 metres at most.

When the ultra-short waves began to be commercially used in recent years, the choice of a suitable receiver was found surprisingly difficult. Simple detectors, with or without high-frequency and low-frequency amplification, proved insufficiently sensitive. They were also very difficult to handle at such high frequencies, where the smallest stray capacity or mechanical vibration can render a signal unreadable. The selectivity of such receivers may actually be too high, because many ultra-short-wave transmitters are not crystal-controlled as yet and show traces of wavelength drift and frequency modulation which make them unreadable on a selective set. The superheterodyne receiver, which we shall study later in this chapter, is capable of excellent performance on all wavelengths, including the shortest, but is always a complex type. Moreover, since its efficiency falls off at very short wavelengths, owing mainly to the increasing losses and inefficiency of the individual valves and components, it becomes necessary to extend the receiver even further if it is to remain sensitive. The superheterodyne is therefore a somewhat elaborate, costly, and bulky receiver when designed for ultra-short-wave use; and though efficient it does not fit in well when simplicity is desired.

The requirements of this special case have been admirably met by the super-regenerative receiver, which gives a simple and cheap set of good sensitivity and is not too selective to accept imperfect transmissions. No doubt it will be increasingly displaced by the superheterodyne as development proceeds, particularly in the case of fixed installations. But in the portable field the super-regenerative circuit is likely to remain invaluable.

Its small size and weight, combined with high efficiency, have made possible the extensive development of mobile radio on ultra-short wavelengths. Examples of this are the American radio-equipped police cars and motor-cycles, which can maintain touch with head-quarters up to some 10 miles, using an aerial which may be only from 2 to 7 feet long. Portable military services for field use in war-time, aeroplane communications, and the police and fire-brigade services of Great Britain

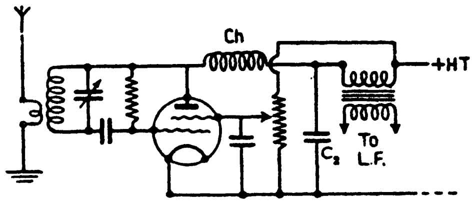


FIG. 109. Modern Self-quenched Detector.

are only a few examples in which the Armstrong circuit has made true portability possible.

The modern super-regenerative circuit is very similar to that of Fig. 107, which has been drawn to illustrate current practice rather than Armstrong's original design. He frequently employed a separate valve as 'regenerator' and another as detector, but this is never found necessary to-day. Also the type of circuit shown in Fig. 108 has passed out of use, being now only of theoretical interest. It is quite possible to combine the regenerative detector function with that of quenching oscillator in a single valve, and this is sometimes done to reduce weight or battery consumption. Fig. 109 shows an example of the circuit used. Here the signal-frequency circuit is of the ultraudion type, specially suited to work on the wavelengths below 10 metres, and is typical of current practice. The valve used is a screen-grid or pentode, which aids sensitivity by its higher amplification factor and allows control of regeneration by adjustment of the screen-grid potential and thus of the amplification factor and R_a of the valve. By joining the grid leak to positive high tension and adjusting its value, the valve is thrown into a 'squeging' type of oscillation due to intermittent charge and discharge of the grid condenser at a frequency determined by the time constant of the leak-condenser combination and by the condenser C_2. This will have a peaky wave form

very suitable to a quenching oscillation, and characterized by long build-up periods each followed by a rapid and complete damping-out of regeneration. The system is thus self-quenching by the 'squeging' oscillation and is practically as efficient as the separately quenched type, but is more difficult to adjust. It is necessary for the valve to regenerate at two frequencies simultaneously, and this requires a nice balance of the various components, such that the two amplitudes will be suitable. This is not easily maintained over any considerable band of wavelengths, the valve tending to oscillate at one frequency or the other only. A separate quenching oscillator is therefore preferable when bulk permits. A low-frequency amplifying stage may be added if necessary, such as when a loud telephone signal is needed in noisy surroundings; but good signals can be received without it. A two- or three-stage amplifier can be used without special precautions when a loud speaker is to be used. Also it is desirable when possible to precede the detector by a high-frequency stage. This serves mainly to prevent radiation of the quenching frequency from the aerial, which can seriously upset neighbouring receivers. It also simplifies adjustment, improves the signal-to-noise ratio, and, if a modern low-capacity valve such as the acorn type be used, may give a small increase in sensitivity. A modern receiver working at 5 metres wavelength has been found to give clearly readable signals from a measured input of less than 5 microvolts. A high-frequency stage may considerably improve this.

Before leaving the super-regenerative circuit, one more property remains to be noted. Since regeneration is constantly occurring and can be set up by a very small signal voltage indeed, all noise from either within or without the circuit will be amplified. This means that the receiver when working near maximum sensitivity always produces a roaring sound in the telephones, more or less intense, and due to the aggregate of all such amplified noises. This is distinctive of the circuit and often termed super-regenerative mush. It would be a serious matter were it not for the fact that the arrival of a true signal of slightly greater amplitude immediately 'takes charge' of the regeneration, and the mush noise is reduced. Strong signals completely remove all traces of mush, whilst weaker ones reduce it to an extent depending upon their strength, which is thus approximately indicated. Unlike most circuits, the super-

regenerative thus receives maximum noise in the absence of a signal, but the arrival of a carrier wave reduces this to about the same level as it would appear on an ordinary receiver, the final result being similar. The mush can be looked upon as a sign that the circuit is in order, rather than that it is fundamentally noisy, and is a sign of high sensitivity. The receiver is found very impervious to certain types of interference prevalent on ultra-short waves, such as car-ignition noise, no doubt because its inherent automatic volume control action limits this to a low value. It is interesting to note that the Post Office engineers, who use both super-regenerative and superheterodyne receivers for their regular telephone radio-links to such points as the Channel Islands, have measured the difference in noise level. This differs little for signals of fair strength, but is some 20 decibels in favour of the superheterodyne in the case of very weak signals.

Published descriptions of portable radio-equipment, often contain references to the 'transceiver'. This is simply a device in which a low-power transmitter and a receiver are combined into a single instrument. Whilst any combination of transmitter and receiver might be so described, the term is properly restricted to a design in which all or some of the same valves and components are used for both transmission and reception on a common wavelength. The use of a common circuit as far as possible naturally saves bulk and weight, the corresponding limitations being inability to send and receive simultaneously or on widely different wavelengths. A single multi-contact switch effects all necessary circuit changes, converting the instrument from transmitter to receiver with a single movement. Transceivers are most used in ultra-short-wave mobile working, and in the receiving position will generally be super-regenerative. The circuit of Fig. 107 will be seen to lend itself readily to easy alteration, for as a receiver we can regard the detector as an oscillator which is modulated (and prevented from continuous oscillation) by the choke-coupled quenching oscillator V_2. The latter resembles a modulator in most respects, and, if the quenching coils be switched out by simple short-circuiting and a microphone with its transformer connected to the grid of V_2, it will act as a modulator of V_1 for speech. At the same time contacts on the switch short-circuit the telephones, so that loud speech is not heard by the operator

if he continues to wear them, and the valve V_1 relieved of the quenching effect will oscillate strongly as a simple self-excited transmitter. There need be no readjustment of reaction or of tuning, provided that the remote station uses a similar transceiver on the same wavelength.

Space has been devoted to the super-regenerative circuit on account of its wide utility under present conditions, but it must be remembered that this does not extend to the longer short waves, where a conventional detector followed by low-frequency stages is the only suitable equivalent, for reasons of selectivity. We have seen that the principles by which a triode valve amplifies are the same for all frequencies, and that were it not for stray capacities a circuit such as the resistance-capacity-coupled amplifier could be used at any frequency. When the valve first came into use as an amplifier, it was found simple to obtain good magnification of speech with the help of either transformer or resistance coupling, and early sets employed a valve or crystal detector followed by several low-frequency stages. Applied to radio-frequency amplification, however, the same circuits gave a very poor performance, and their use remained limited for some years. Eventually resonant impedances in the anode circuit were found to overcome this trouble, for the stray capacities became part of the tuned-circuit capacity and lost their objectionable shunting propensities. The 'tuned-anode circuit' became popular, and effective amplification at the frequency of the incoming signals became possible, together with increased selectivity. This greatly extended the range of the simple detector, giving in conjunction with reaction and low-frequency stages a better sensitivity and selectivity than ever before. Stray capacities remained a limitation, however, for they set up instability and self-oscillation in the high-frequency stages, which remained very troublesome and reduced the maximum usable amplification until the development of screen-grid valves finally overcame that particular defect.

It will be interesting to consider the steps taken, before the introduction of the screen-grid valve, to prevent the interelectrode capacities of a triode from setting up a coupling between anode and grid circuits, and thus causing self-oscillation of the stage. Whilst no longer met with in receiver design, large triodes are still important in transmitting work and must often be stabilized by similar methods. Any process by which

this coupling can be prevented or counteracted is termed neutralization, and many varieties of this were in use at one time or another. Most rely for their action on a compensating coupling arranged to transfer equal energy from anode to grid circuit, but in the opposite phase to that transferred by the electrode capacities. Such circuits resemble an alternating-current Wheatstone-bridge network in which the anode circuit represents the source of alternating potential, and the grid circuit the output from the bridge, which becomes zero when balance exists. The principle will be made clear from Fig. 110, which shows the essentials of the most widely used arrangement for neutralization.

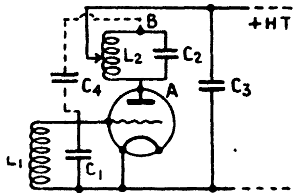


FIG. 110. Neutralization.

In full line is shown a triode amplifier. L_1C_1 forms the tuned-grid circuit and must be assumed to be fed with the signals from an aerial or preceding stage which is not shown. Similarly, the anode circuit contains a tuned circuit L_2C_2, which may be coupled to the following stage in any manner. In the present case it will be noticed that the anode voltage reaches the anode through a tapping on L_2 assumed to be at the mid point of the inductance, and which is 'earthed' to the valve cathode through a condenser C_3 of negligible reactance. Under this condition there will be an equal radio-frequency potential at each end of the coil L_2, but that at A will be in opposite phase to that at B. Now from A there will be a transfer of potential through the internal capacity of anode to grid, which will be in the correct phase to cause reaction and possibly self-oscillation. The problem becomes that of transferring to the grid an equal potential, of opposite phase, able to neutralize the former. Such a potential exists at the other end B of the anode coil. If therefore a condenser C_4 of capacity equal to the anode-grid capacity be connected from B to the grid, as shown in the dotted connexion, an equal potential will be transferred. No energy will now be returned from anode to grid circuit through the valve capacity, and if the circuits are externally screened it should be possible to make stable use of the full amplification of which the valve is capable. By analogy with a bridge circuit, the two halves of L_2 form the ratio arms, whilst if the anode-grid capacity be treated as the unknown capacity, the added condenser C_4

becomes the balancing adjustment. Calling the two halves of L_2 L_a and L_b respectively, then the relationship for neutralization becomes:

$$L_a/L_b = C_4/C_v.$$

It will be noticed that only half of L_2, the portion termed L_a, is actually in the anode circuit. This may imply a loss of amplification, unless the wavelength is long enough for the impedance of the tuned circuit L_2C_2 to exceed about four times the R_a of the valve. Neutralization generally implies some sacrifice of this kind, but by taking advantage of the relationship obtained from the bridge treatment it is possible to minimize that loss. In the case given, $L_a = L_b$, and so C_4 must equal C_v. This means a very small condenser equal to the valve capacity, and such small 'neutralizing condensers' were a common feature of the wireless store a few years ago. More convenient condensers can be used if L_2 be tapped unequally, and at the same time a larger portion of the circuit will be included as useful anode load. Thus, let the tapping be such that $L_a = 10L_b$. Then, from the equation, C_4 will become $10C_v$. A variable condenser of larger capacity can now be used, simplifying adjustment and raising the amplification of the stage by the inclusion of nearly all of L_2 in the anode circuit. Neutralization is the only part of early amplification technique which has a direct application to-day, and it will be unnecessary to look any further into the expedients that were used to squeeze reasonable stable amplification from triodes. All were very inefficient in comparison with a simple screen-grid or pentode stage, which may show an amplification of from 50 to 250, as against 5 to 20 for a neutralized triode stage.

With the introduction of the screen-grid valve, amplification at signal frequency became at once a practical proposition. No steps are usually necessary to preserve stability beyond complete isolation of the external anode and grid circuits. It is hardly necessary to point out that unwanted reaction can easily occur by magnetic or other coupling between the anode and grid circuits, and particularly by interaction between the fields of the respective coils. In the case of modern valves this is usually the limiting factor in stable amplification. Even when insufficient to cause instability, it may reduce sensitivity, impair selectivity, and make the receiver unpleasant to handle. Early receivers having low stage gain sometimes relied upon

mere spacing to reduce these couplings, the sets being bulky and stages well removed from each other. Later, metal screens were introduced, often in the form of plates, and some receivers evolved into elaborate metal honeycombs having a separate compartment for each stage. Whilst quite effective, these were difficult to build and repair. They have gradually given way to a construction that is both simple and effective, consisting of

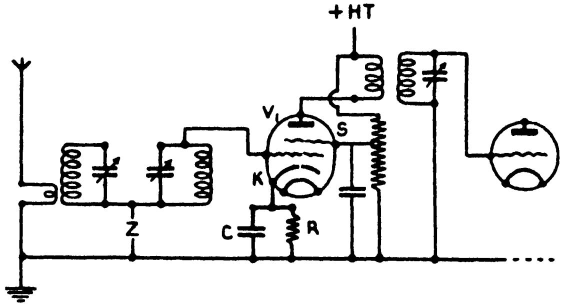


FIG. 111. Screen-Grid Amplifier.

a metal base-plate, or 'chassis', raised to allow of small components and wiring beneath. This replaces the original wooden base-board and often takes the form of a shallow inverted tray of steel or aluminium. All coils and tuned circuits, together in many cases with the valves and other vital components at radio-frequency potential, are enclosed in individual metal containers, which are often cylindrical aluminium cans. It is found that if resonant circuits and those components having an extensive field are individually screened in this way, the remainder of the circuit wiring can be exposed beneath the metal chassis in safety. Capacitive coupling is reduced by the presence of the large earthed metal mass, whilst the screening cans introduce earthed metal between components which might interact.

The circuit of a typical modern screen-grid or pentode stage is shown in Fig. 111. It will be seen to be quite simple, relying for efficiency upon the design of the valve and components and upon thorough screening and sound layout. Should this stage precede the detector of a normal three- or four-stage receiver, it would probably be preceded by two tuned circuits loosely coupled to improve selectivity. These would be individually screened, and coupled by the reactance Z common to both, which will be discussed later. The aerial will be loosely coupled

to the first circuit, whilst the second is joined between grid and earth. A small negative bias is necessary for the correct functioning of the valve, and this can be obtained in the case of indirectly heated valves by the system termed 'cathode biasing'. A resistance of a few hundred ohms is inserted at R, between the valve cathode and the negative high-tension or earth line. The whole anode current of the valve will flow through this resistance, producing across it a potential drop which provides bias. From the circuit it will be seen that the valve cathode becomes positive by this amount with respect to the negative high tension, and since the grid is joined to the latter through its tuning inductance (or on occasions a grid leak) it must become negative with respect to the cathode. The correct value of cathode-biasing resistance is given simply by Ohm's Law. If I_a be the desired working anode current of the valve (which must include the screen-grid current when this is appreciable) and V_g the desired working grid bias corresponding to that anode current, then the resistance R is given by the ratio V_g/I_a in ohms. The condenser C must be joined across the biasing resistance, so that the cathode is substantially at earth potential for radio-frequency, otherwise the resistance, being common to both grid and anode return circuits, would be a source of coupling between them. A mica condenser of from 0·01 to 0·1 microfarad is suitable at broadcast frequencies, but when the biasing scheme is used in low-frequency stages it must be of much larger capacity.

The screen-grid potential is best obtained from a potential-divider across the high-tension supply as shown, the resistance value depending upon the voltage needed by the valve. It is necessary here also for a by-pass condenser of negligible reactance to be joined from screen-grid to cathode, so that the screen may not take up any radio-frequency potential which might upset the correct working of the stage. In the anode circuit we have a choice between a single-tuned circuit, a transformer with untuned primary winding as shown, or a loosely coupled transformer having both primary and secondary tuned individually. We have reviewed the need for these in an earlier chapter, and shall consider them from the point of view of selectivity later. Meanwhile there are practical considerations which may determine the type of coupling. A single-tuned circuit will be convenient when several wave bands have to be covered by

inductance switching, as the switch design will be simplified. It makes for cheap and easy construction. The transformer with untuned primary will be a little more complex, particularly if it is to be switched. On the other hand it may provide slightly better amplification and selectivity, particularly if the valve be a high-resistance type, whilst the primary provides an ideal path for the anode current, entirely insulated from the secondary. This is said to prevent noise on short waves, which can be caused by slight leakage across a coupling condenser. The double-tuned transformer will give better selectivity and is usually selected if a fixed wavelength is to be amplified, as in the intermediate-frequency stages of a superheterodyne receiver. The need for two variable condensers and added switching would usually tell against it in the case of a variably tuned amplifier.

Designed on these lines a signal-frequency stage will amplify well over long and medium wavelengths. The gain will naturally depend upon how nearly the R_a of the valve can be approached by the impedance of the anode load, which we have seen should be not less than twice R_a. Consider the simplest case, in which the anode circuit contains a parallel resonant circuit. The impedance of this circuit is equivalent to a pure resistance at resonance, and its value is given by $Z_a = L/RC$ ohms. Now resonance occurs at the wavelength given by $\lambda = 2\pi LC$, and so the product LC must fall as the wavelength is reduced. It has been pointed out that for practical reasons C cannot be reduced indefinitely, and so both L and C must be reduced together if a very low wavelength is required. Z_a will be a maximum when the ratio L/C is a maximum, but at low wavelengths, where C is already a minimum determined by the stray circuit and valve capacities, further reduction is only possible by reducing L. This implies an unavoidable reduction of the impedance Z_a also. Practically, C must not be too low if reasonable selectivity is desired, thus still further limiting the L/C ratio. Also the term R, which includes all circuit losses, will increase in proportion to L as wavelength is reduced, because of the poorer properties of dielectrics and the greater resistance of conductors at high frequencies. This factor will still further reduce the impedance Z_a.

The combined effect of all these factors is that Z_a must fall with wavelength to a greater or lesser extent. At medium wave-

lengths above perhaps 100 metres it will be possible for Z_a to exceed $2R_a$ for a normal screen-grid valve; but below this figure Z_a unavoidably falls, and at the shortest wavelengths it may be only a few per cent. of R_a. Amplification must therefore suffer, and modern development in high-frequency-stage design has been mainly directed to the minimizing of this loss. Three main lines of attack are open. The one is to reduce all stray capacities to an absolute minimum, such as by the careful design of components, short leads and small dimensions throughout, and small valves, such as the 'acorn', having minimum inter-electrode capacities. The second is the design of all components to have minimum losses at high frequencies, mainly by the use of specially efficient synthetic dielectric materials, and highly conductive leads and coils, which may be silver-plated. The first of these keeps the L/C ratio as high as possible, whilst the second reduces all the factors contained in R to a minimum. Thirdly, some improvement is possible by better valve design. In the case where Z_a is less than R_a, the ratio Z_a/R_a is practically a measure of the proportion of the valve's amplification that can be used. Thus a valve having low R_a and maximum amplification factor μ is desirable, which means a valve of the highest possible mutual conductance. Now the mutual conductance of valves varies from one type to another far less than does R_a or μ, and so we find that valves differ less than might be expected in their performance at low wavelengths. The increase of mutual conductance of modern valves from less than unity to anything between 3 and 10 has correspondingly increased the amplification possible. Practically, however, the best valves tend to be those having minimum internal capacity, a low R_a permitting of reasonable impedance matching, and the highest mutual conductance obtainable under those conditions.

Having obtained the most efficient valves and coupling impedance possible, the signal-frequency amplifier for very short waves differs little from its medium-wave counterpart. By-pass condensers and circuit layout must be more carefully watched, since a smaller stray capacity or common impedance can effectively couple two circuits, and instability is very easily caused. By-pass condensers particularly must be non-inductively constructed and have the minimum possible power factor, mica dielectric rather than paper being essential. The metal 'chassis-

and-can' type of assembly becomes less satisfactory at the shorter wavelengths, and it has been shown that high-frequency eddy currents flowing through a small area of metal common to the screening of two circuits can introduce appreciable coupling. A different construction is found preferable in advanced short-wave receivers, in which a complete metal screening box or compartment includes each individual stage, with its associated tuned circuit. These boxes must have no common walls, and if in close proximity must be insulated from each other at all but one point, where they are bonded to a common earth line. Also circuits must not be earthed to the boxes or chassis at various points, treating this as a large metal body at zero potential, as is permissible at longer wavelengths. All circuits must be returned to the earth line or negative high tension by separate insulated conductors, and earthed at one common point only. Provided these precautions are adhered to, screening should present no difficulties even in the ultra-short-wave region. Owing to the lower stage amplification at these wavelengths, and provided that not more than one stage works at any particular frequency, complete screening is not always essential. Circuits may sometimes be used well spaced out on an insulating board, a construction often possible for transmitters. Discretion must be used in this connexion, and experiment is usually the best way to determine whether any given system of screening is adequate.

The high-frequency coupling transformer comprising two separately tuned windings will not in general yield any increased amplification from the stages which it couples. Provided that weak coupling is employed, amplification will be lower than that of a tuned-anode circuit, by an amount depending upon the looseness of coupling. Selectivity will, however, be increased to the product of that of the two individual circuits, being squared if they are identical. It is usual to employ similar L/C ratio for each circuit, although some improvement may result by raising this ratio somewhat for the secondary. As coupling is increased, the amplification will also increase until optimum coupling is reached, when the amplification will be of the same order as for the tuned-anode arrangement. Before this point is reached, however, the selectivity curve will tend to become double-humped, and selectivity will be reduced. It is not in general possible to obtain maximum selectivity and amplifica-

tion for the same coefficient of coupling, and a compromise must be arrived at.

A transformer having untuned and closely coupled windings will behave more as would a low-frequency transformer, and can be somewhat similarly treated. An auto-transformer can also be realized, if either the valve anode or the following grid be tapped down so that only a portion of the whole inductance is in circuit. This is often advantageous, since it reduces the load across the resonant circuit and thus improves its selectivity by sharpening the resonance curve. The full calculation of transformer ratios is beyond the scope of the present book but is approached from considerations of impedance matching. A variety of conflicting factors have to be combined to produce the optimum compromise. Assuming maximum amplification is aimed at, rather than selectivity, it is first necessary to ascertain either by calculation or measurement the resonant impedance of the secondary winding, which should be the highest obtainable. Should this be less than about $2R_a$, it will be entirely included in the anode circuit; but otherwise a primary winding or tapping is selected such that the ratio of primary to secondary turns equals the square root of the ratio of R_a to total impedance. In estimating the total or secondary impedance it may be necessary to allow for the load placed across it by the grid-cathode circuit of the following valve. At longer wavelengths this may be negligible, but at shorter wavelengths it becomes comparatively large. It is now necessary to match this impedance also, either by selecting a tapping-point for the grid lead or by adjusting the number of turns and L/C ratio of the secondary. The final relationship aimed at is that in which the square of the turns ratio between primary and secondary equals the ratio of R_a to the impedance of the following grid circuit, when energy transfer will be a maximum. Other considerations may, however, prevent this condition being exactly attained. Thus at short waves the necessary inductance values may be too large to resonate at the desired frequency. Also, since the transformer is usually to be tuned over a band of wavelengths of perhaps 3:1, the ratio cannot be correct over the whole range. This factor makes the selection of ratio somewhat approximate and not generally critical, the ratio being nearly 1:1 for valves of high R_a and becoming somewhat higher (lower primary inductance) as the wavelength or valve resistance decreases.

Earlier broadcast receivers and many present commercial types employ separately adjusted tuning condensers for each circuit used to couple the high-frequency stages or detector. This may imply three, four, or more tuning controls, and, whilst not necessarily objectionable in commercial installations, where a receiver may remain tuned to a single transmitter for long periods by a skilled operator, it is a serious tax upon the ability of the public. As a result, attempts have been made for many years to develop a single-tuning-control receiver, by the process of 'ganging' all variable condensers upon a common spindle. For some time such receivers were not entirely satisfactory, as it was not found possible to maintain all circuits accurately tuned to the same frequency over a wide wavelength range. Steady commercial development has gradually overcome this difficulty, until at the present time most receivers employ a single ganged control, with usually only a very slight loss of efficiency. Two lines of development have made this possible: the one that of closer manufacturing tolerances and greater precision in the variable condensers and components themselves; and the second, minor circuit modifications which render the tuning less dependent upon variable factors such as aerial loading or valve characteristics. The sections of a ganged variable condenser can now be matched to within 0·5 per cent. under mass-production conditions. This still results in slight loss of efficiency where several high-frequency stages are to be ganged, but we shall see that it is quite sufficient for superheterodyne purposes. As regards circuit coupling, the increasing efficiency of modern valves has made it possible to work with looser couplings throughout than was once usual. Thus the aerial can be coupled so loosely that it has little effect upon the tuning of the initial circuit, and similarly intervalve couplings can be loosened when necessary, thus achieving improved independence of tuning. The loss in amplification which results is more than made up by increased valve efficiency.

We shall look further into these matters when discussing selectivity in a following chapter. Meanwhile, we can sum up the outstanding characteristics of a modern receiver employing one or more stages of high-frequency amplification tuned to the incoming signal, a regenerative detector, and any desired degree of low-frequency amplification. Such a receiver is of course a great advance on earlier types, in which no amplification

preceded the detector, even when the latter was regenerative. It opens up the possibility of using any preferred detecting system, such as a diode when high-quality reproduction is required, and of making up for lost sensitivity by increased high-frequency amplification. The regenerative grid detector is only used with a single high-frequency stage, for when preceded by two or more stages it becomes difficult to stabilize. The sensitivity of such receivers is good, a typical figure being 50 microvolts at the first grid for a standard output of good audibility, taken as 50 milliwatts of audio power from a carrier wave modulated 30 per cent. by a 400-cycle tone. This is adequate for many uses, but can be exceeded by the superheterodyne, where the figure may be better than 1 microvolt. It is possible with the straight receiver to hear the weakest useful signals with a trained ear, but not always easy to obtain sufficient amplification for their use in automatic reception. In selectivity the receivers vary considerably with the number of tuned stages incorporated, but, whilst they may be adequate for the separation of relatively weak signals at less than 10 kilocycles frequency difference, they are not able to separate strong or local signals under similar conditions. We shall see that this selectivity is not sufficient for all commercial purposes.

The need for a better receiver is best realized, however, by noting the deficiencies of the straight type. As just stated, the sensitivity is not always adequate for commercial uses. More important is the lack of selectivity, which implies a somewhat elaborate chain of ganged circuits if it is to be good, and at its best will not prevent interference from powerful transmitters. In addition it will be shown that straight circuits do not lend themselves well to various recent circuit improvements described at the end of this chapter. In early years these defects were still more noticeable and gave impetus to the development of the superheterodyne circuit now to be considered. The principles of this have been known for a considerable time, but since in its crude forms the circuit bristles with difficulties and is inherently costly its development was slow. Ten years ago the superheterodyne was termed 'the Rolls-Royce of receivers' and found only in leading commercial installations and a few private homes, although it may be noted that in France and some other countries crude forms of it found greater popularity. The salient feature of this system is that selectivity can be very

high, and an ever larger number of transmitters has brought it into great prominence. Modern valve and circuit improvements have gradually eliminated most of the defects found in early superheterodynes and, with the help of valve types evolved solely for that purpose, have gradually raised it to its present position as the leading system of reception.

The essential superheterodyne principle is illustrated in

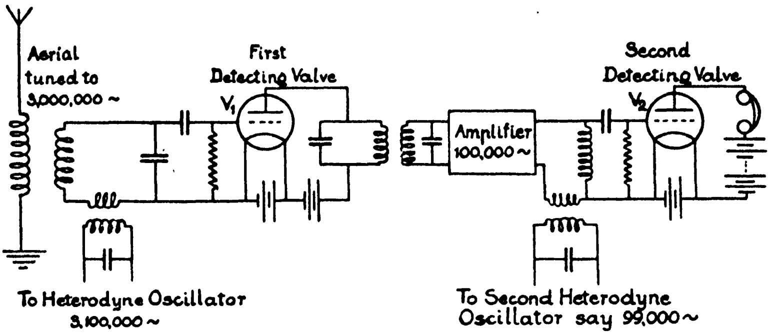


FIG. 112. The Superheterodyne.

Fig. 112. The incoming high-frequency signal is heterodyned by a local oscillator in the way previously explained, but adjusted to a frequency which will produce super-audible beats of a definite frequency termed the 'intermediate frequency' (abbreviated to I.F.). In the diagram the signal frequency is taken as 3 megacycles, and the local oscillator as 3,100,000 cycles, which will produce beats at the difference between these two, namely 100,000 cycles per second. The intermediate frequency is thus 100 kilocycles in this example, but it may be chosen within wide limits in the long-radio-wave region. It should be noticed that a local oscillation of 2,900,000 will produce the *same* beat frequency of 100 Kc., and that there are thus always two frequencies of local oscillation that can be used to heterodyne a given signal and produce a given intermediate frequency. The one of these actually used is called the 'main-channel' adjustment, and the other the 'second-channel' adjustment. Expressing this in general terms, a signal of frequency X can be heterodyned to produce an I.F. of F Kc. by a local oscillator set at frequencies of either $(X-F)$ or $(X+F)$. Extending this, it will be seen that for a given oscillator setting of, say, Y Kc., the I.F. of F Kc. can be produced by *two* incoming signals, namely

at $(Y-F)$ and $(Y+F)$ Kc. If therefore one of these is in use, signals at the second frequency might be received at the same time unless tuned circuits are introduced to keep them out. Interference of this kind is termed 'second-channel' interference. Circuits following the aerial and tuned to the wanted signals for the main purpose of preventing this are termed 'preselecting' circuits.

Reverting to the basic principle, the selected I.F. is produced by heterodyning as explained, and this process must occur in a detector valve V_1 in the illustration. This was originally termed the 'first-detector stage', but is now more often termed the 'mixer stage' or simply the 'frequency-changer', being that in which the incoming signals are mixed with the local oscillator. Clearly, by the correct oscillator setting any incoming signal can be heterodyned to the desired I.F., in this case 100 Kc. Following the mixer, therefore, will be an amplifier designed and adjusted to work only at the I.F. of 100 Kc. This is termed the 'intermediate-frequency amplifier', and will employ high-frequency stages working efficiently at that frequency. After amplification, the signals, still at the super-audible frequency of 100 Kc., must be rectified again by a 'second detector' V_2. They then become identical with those normally produced by the detection of a 100 Kc. radio-signal, and are heard in the telephones or passed to a low-frequency amplifier as usual. The second local oscillator, shown at 99 Kc., is optional, serving merely to heterodyne continuous-wave telegraphic signals to an audible note, in this case 1,000 cycles. When receiving modulated waves such as telephony it is unnecessary.

The beauty of superheterodyne reception lies in the use of a fixed intermediate frequency at which good amplification is possible, and thus the I.F. amplifier may contain as many stages as desired and be of very high gain indeed. It is easily possible to amplify in this way up to the limit set by fundamental valve and circuit noises, and one can say that the sensitivity possible is the maximum which our knowledge of physics will allow. Still more important is the fact that in the I.F. amplifier any number of resonant circuits can be used, and, since their tuning is fixed to one frequency, little difficulty exists in maintaining precise adjustment. Selectivity can therefore be built into this amplifier to an almost unlimited extent, determined by that which is desirable rather than that obtainable. The use of a fixed

frequency naturally facilitates a wide variety of circuit devices which would be too complex for satisfactory use in circuits retuned for every incoming signal.

It will now be best to review the various sections which make up a complete superheterodyne receiver, bearing in mind that they have passed through innumerable modifications, and that space will only allow mention of the best modern practices, with brief reference to those which have preceded them. Firstly, it is clearly possible to amplify the incoming signals before they are heterodyned to the intermediate frequency, using for the purpose high-frequency stages exactly similar to those already described. Such amplification is less necessary in the superheterodyne, because considerable amplification is now available in the I.F. amplifier; but it may be advantageous for the dual purpose of raising the signals above valve noises occurring in the frequency-changing circuits, and to provide the tuned circuits necessary for preselection.

The introduction of tuned stages will naturally complicate the adjustment of the receiver. If these are ganged to the oscillator for single-dial control they introduce a new ganging problem, because the oscillator must be on a different frequency from the signals and maintained at a constant frequency-difference from them. The problem is less acute, however, because the tuning of the receiver as a whole is determined by the oscillator setting only, a given signal being received only when the oscillator frequency is such as to heterodyne it to the exact intermediate frequency. The I.F. amplifier will be by far the most selective section of the receiver, and need not respond to signals differing by more than a few kilocycles from the correct frequency. The resonance of the preselecting circuits will be relatively flat, and if not always exactly correct will only reduce the strength of signals somewhat. Unlike the straight receiver, a small inaccuracy in the tuning of the high-frequency stages will have very little effect upon selectivity.

In early superheterodyne design the necessary frequency-difference between oscillator and signal circuits was maintained by the use of different inductance values, combined with series or parallel 'padding' condensers which modified the value of the main oscillator-tuning condenser. At that time similar ganged condensers served for each stage, and padding was necessary to obtain different capacity values in the oscillator circuit. The

ideal is clearly a different rate of change of capacity with dial setting for the oscillator, and this is obtained to-day by the use of a section of the ganged condenser assembly having specially shaped vanes. The contour of these vanes can be designed so that a constant-frequency difference will be maintained between the oscillator and signal-frequency circuits, thus overcoming the main difficulty which used to prevent perfectly ganged super-heterodyne receivers. Quite a number of commercial receivers may be met with, however, in which the older padding condensers are used, mainly to save the cost of a specially designed ganged condenser.

This will be a convenient juncture at which to refer briefly to the changes which have taken place in the design of variable condensers. As originally evolved, these invariably possessed semi-circular vanes, and since capacity is proportional to the area of overlap between the two sets of vanes the law connecting capacity with dial rotation was a linear one. Neglecting 'end effects' due to the edges of the vanes, which modify the law slightly at the extremes of the scale, the relationship would be expressed by

$$C = K\Delta+\beta,$$

where C represents the capacity of the condenser for an angular displacement Δ from the minimum position, K a constant depending upon the dimensions of the condenser, and β the residual capacity at zero angular displacement, or the 'zero capacity' of the device. Now from the expression $\lambda = 2\pi\sqrt{(LC)}$, which can be written $\lambda^2 = 4\pi^2LC$, we see that when L is constant λ^2 varies as C. Thus, for the semicircular-vane condenser tuning a given fixed inductance, wavelength will be proportional to the square root of the dial rotation. This is not an ideal condition, since it must lead to a crowding of stations near the zero end of the condenser scale, where the rate of change of capacity is large; whilst at the maximum end of the scale large angular changes will produce much smaller changes in wavelength. Both to avoid crowding of the scale and to facilitate calibration, it will be better to have a condenser in which wavelength will be directly proportional to dial rotation, rather than to the square root of this. Such a condenser can be produced if the vanes be of varying radius, instead of semicircular. It will have a law of capacity variation given by

$$C = K\Delta^2+\beta,$$

and for this reason is called a square-law condenser. The general adoption of square-law construction much improved ease of adjustment and calibration of receivers, rendering the old semicircular type obsolete except for laboratory purposes, where a linear law may sometimes be convenient.

With the introduction of shaped vanes to give a square-law capacity variation, makers realized that other improvements along similar lines would be possible, and a variety of more complex types made their appearance. By stating the law required to yield any particular tuning curve in the form of an equation, and either by solving this graphically for a large number of progressively increasing angular displacements or by solving the corresponding differential equation for the locus which corresponds to the outline of the vanes, a condenser can be designed to comply with any conditions. An important case is that which, tuning a specified inductance, will give a straight line when dial setting is plotted against frequency. It was first introduced under the name of 'Kc' condenser by the Dubilier Company. Once it had been realized that the spacing of transmissions along a receiver scale was fundamentally a matter of frequency rather than wavelength, such a condenser became the obvious choice, for it resulted in a scale on which transmissions separated by equal frequency-differences would appear at equal increments of rotation. This is the most uniform arrangement possible, being independent of actual wavelength. Other designs were produced in which the law was modified to produce real or imaginary advantages, and in which the effects of self-capacity within the inductances were allowed for.

Now, if it is possible to shape condenser vanes so that a linear law of rotation against frequency results, it is equally possible to design two condensers in which the frequency term differs by a constant factor, such that the two curves run parallel at a definite frequency-difference. This will merely imply vanes of the same shape, giving equal rates of change of frequency with rotation, but differing in actual area, so that a constant frequency-difference exists. Two or more such condensers ganged upon a common spindle will solve the ganging problem for the superheterodyne.

Considered from the point of view of signal-frequency amplification, superheterodynes can be divided into two classes, namely those in which this is fully made use of and those in

which it is kept down to a minimum or entirely dispensed with. Receivers for commercial communication normally typify the first of these classes, employing considerable amplification before the frequency is changed. Such receivers are often tuned to some particular transmitter for very long periods, in which case individual tuning controls can be used for each circuit, and the delay caused by the resetting of several controls is not important. In this case additional tuned stages offer no particular difficulties and may be used freely. The main objection to them in receivers designed for frequent change from station to station is a considerable complication in construction, owing to the several circuits which must be ganged or adjusted for each change. Wave-change switching is also elaborated.

Radio-frequency amplification is unquestionably desirable in superheterodynes intended for high sensitivity and long-range working. It has been mentioned that a deciding factor in such receivers is the ratio of signal to noise generated within the circuits, and this is somewhat higher in the case of the frequency-changing circuits than of a normal amplifying valve. The first detector will be carrying two components of oscillatory anode current, one due to the incoming signals and the other from the local oscillator. These combine to produce a signal at the intermediate frequency, and this signal will be modulated by any noise produced in the detector or the oscillator. For very weak signals the latter may predominate and will introduce a proportion of 'valve hiss' which may be large in comparison with the signal. It is thus an advantage to amplify the signal first, so that it may be as strong as possible at the first-detector grid and able to overcome noise introduced at that point. There is a limit to the amplification desirable at intermediate frequency, for this same reason, that slight unavoidable valve noise from the frequency-changer will be magnified until it becomes objectionable. Thus if more amplification is wanted it must be introduced before frequency-changing.

It has been shown that a maximum of two stages of signal-frequency amplification is ample to ensure the best possible noise ratio. The first valve in these stages must unfortunately introduce some noise of its own, owing to Miller effect, and this will lie between one-fifth and one-twentieth of that likely to occur in the first detector. Even a single stage can amplify more than 20 times at most wavelengths, and can therefore amplify

the signals to an extent where noise from the first valve becomes the limiting factor. This limit lies somewhere near 5 microvolts oscillatory potential at the first grid, corresponding to a field strength which may be well below 1 microvolt per metre of aerial height. Two stages are more often used when cost permits, because they ensure sufficient amplification over the whole wave range of the receiver, including possible bad spots where efficiency is lost through switching, or at the shortest wavelengths. This will imply at least three tuned-coupling circuits, giving ample preselectivity for the elimination of second-channel interference and allowing a low intermediate frequency to be used, which we shall see is an advantage from the point of view of selectivity. Advanced commercial receivers may employ more than two stages, but mainly for the purpose of still more complete preselection; whilst many receivers, including most of the better-class broadcast designs, obtain excellent results with a single stage only.

On the other hand there are many occasions when very high sensitivity is quite unnecessary, and signals so weak as to compare with the valve noises are not wanted. This condition occurs in the bulk of less expensive broadcast receivers and commercial circuits working under good conditions or over short ranges. Signal-frequency stages are then quite unnecessary, since the signals will be strong enough for direct application to the frequency-changer. If they be omitted, however, a new problem arises in preselection, since only a single-tuned aerial circuit or possibly a pair of loose-coupled circuits would be used at signal frequency. These may not attenuate second-channel interference sufficiently, particularly if the I.F. be low. Interference is to be expected from stations removed by twice the I.F. from those being received, and the preselecting circuits must be selective enough to reject signals at that frequency off resonance. Thus, for an I.F. of 100 Kc. the circuits must be able to reject signals 200 Kc. off resonance, their response at that frequency being perhaps 10,000 times less than at resonance. At broadcast frequencies of some 1,000 Kc. this represents 20 per cent., whilst at short waves it may be only 2 or 0·2 per cent. In the latter case a pair of averagely damped circuits would be quite unable to effect the separation, and every signal would be tuneable at two points on the dial, a fault often found in poor-class receivers.

Ingenious methods have been evolved to reduce this difficulty. Much can be done if the resonance of the preselecting circuits be sharpened by reaction, but at the expense of critical adjustment and possible ganging defects. An ingenious American scheme employs reaction to effect this, arranging the reaction coupling to operate in reversed phase at the second-channel frequency, when any incoming interference will be reduced in amplitude by 'negative regeneration'. Excellent results are claimed. By far the most popular method, however, lies simply in the use of a higher intermediate frequency, a very general figure being 465 Kc. This will shift second-channel interference to some 930 Kc. from resonance, and quite moderate preselectivity can then reject it. 465 Kc. lies outside both European broadcast bands, and if the oscillator be set to the high-frequency side of this, second-channel interference can only come from stations on still higher frequencies. It then becomes possible to fit low-pass filters between the aerial and first detector, which will pass the signal frequencies whilst rejecting the higher second-channel range. Commercial receivers are not limited by the requirements of broadcast wavelengths, and can employ a wide selection of higher intermediate frequencies. These are chosen for a variety of reasons depending upon the service for which the receiver is to be used, the main consideration being most often selectivity, to be discussed in the next chapter. Receivers intended only for short-wave reception may employ still higher I.F., of which 1,600 Kc. is an example. Ultra-short-wave receivers and those used for television may use frequencies as high as from 3 to 20 megacycles, mainly to obtain the desired wide band-pass characteristic in the I.F. amplifier.

At one time superheterodynes in which an intermediate frequency higher than the incoming signal was used were termed 'infradyne' receivers; but since there is no fundamental difference in their operation, except that second-channel elimination is made easier, no distinction is usually made to-day. It will be shown presently that advantages sometimes accrue from double superheterodyning, or changing the frequency of incoming signals twice, to two different intermediate frequencies. This is perfectly possible in the case of commercial receivers, in which complexity is subordinate to the best possible performance. As an example of the use of this double frequency change, signals may first be heterodyned to a high I.F., which

will make second-channel elimination very complete even at short wavelengths, after which a second change is made to a low I.F., at which high selectivity can be obtained.

We must now pass on to consider the frequency-changing circuits. These comprise two essential parts, the first-detector stage and the local oscillator by which the signal is heterodyned. An early and very simple circuit is termed the 'autodyne', in which these two functions are combined into a single oscillating detector valve, exactly of the type discussed under autodyne receivers. This is only efficient when a low I.F. can be used, because the grid circuit of this valve is compelled to work at two frequencies simultaneously. It must be tuned to the frequency at which oscillation is taking place, whilst it must also offer a high impedance to the incoming signals, which differ by the intermediate frequency. Clearly this compromise will only be satisfactory when the frequency difference is small, for if it be large the grid circuit will be so far out of tune for the signals that very little potential will reach the grid. A low I.F., on the other hand, makes preselection very difficult, and we are faced with the fact that an efficient autodyne arrangement, whilst sensitive, will generally be subject to considerable second-channel interference. It is thus rather crude, and in the original triode form is almost obsolete; but it should be noticed that some recent systems employing a single multiple valve show increasing resemblance to it.

If a separate oscillator be used, the first detector may follow very closely the principles laid down when detection was discussed, a favourite arrangement being a pentode employing grid leak and condenser. The main requirement is efficient rectification. The oscillator may employ a triode in one of the circuits given in Fig. 58, or possibly a pentode in some form of cathode-coupled circuit. The main requirements are a ready oscillation of sufficient amplitude, which may lie between 5 and 30 volts R.M.S. oscillatory potential, combined with good frequency stability. The oscillator must not 'drift' unduly as the valve becomes warm, or be subject to frequency changes if the supply voltages vary somewhat, as such changes will throw a signal out of resonance with the selective I.F. amplifier. The circuit of Fig. 113 is typical of current practice, having excellent stability. If the output be drawn from the anode circuit at *A* the circuit is termed an 'electron-coupled oscillator', because

the anode load is only coupled to the oscillating inner electrodes through the electron stream. This has the advantage that changes in load cannot be thrown back in such as manner as to affect frequency stability.

It is in the matter of coupling between oscillator and detector that superheterodyne circuits have differed most widely, almost

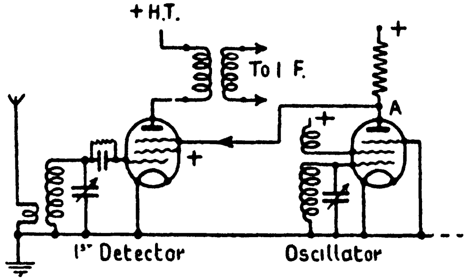


FIG. 113. The Frequency-changer.

every possible form having been advocated. The most obvious are: capacitive coupling through a small condenser between oscillator grid or anode and detector grid; inductive coupling between the respective grid circuits, or into a coil of a few turns introduced into the detector-grid circuit. Should the detector be a screen-grid or pentode type, one of the outer grids may be joined directly to the grid of the oscillator, with good results. This is termed 'grid injection'.

Broadly speaking it is immaterial how coupling is effected, provided that the electron stream within the detector is sufficiently modulated by oscillatory potentials from the oscillator, whilst at the same time a minimum of energy is transferred in the reverse direction. The latter requirement partly decides the merit of a circuit, since interaction by the detector or preselecting circuits upon the oscillator leads to inefficient operation and tricky adjustment, particularly when selectivity is high and wavelengths short.

The most perfect method, adopted in many high-class commercial receivers, is to employ a 'buffer' valve between oscillator and detector. This is merely a screened amplifier, which passes the energy in the required direction whilst being virtually non-conductive in the other. The next-best arrangement has been proved by experience to be electron coupling, as shown in Fig. 113, and is gradually displacing earlier systems. Its effec-

tiveness has led to the development of combined oscillator and first-detector valves, in which a common cathode and electron stream serves for both duties. These may be aptly described as mixing or frequency-changing valves. They are given many names, such as octode, pentagrid, heptode, triode pentode, and triode hexode; but, whilst differing in the number and arrangement of electrodes, their fundamental action is very similar. Next the cathode of the octode are two grids, which function as the grid and anode of a normal triode oscillator connected to suitable external circuits. Since, however, the 'anode' of this arrangement is actually a grid, the electron stream will pass partly through it to the outer electrodes. The electron stream will, however, have been modulated by the oscillatory potential of these inner electrodes, and so the space immediately beyond the second grid can be regarded as a 'virtual cathode' emitting a modulated electron stream. The outer electrodes are connected to form a detecting combination, such as a pentode. Signals will be applied to one of the outer grids, whilst the final anode will receive the combined heterodyne potentials for transfer to the I.F. amplifier. This arrangement clearly resembles electron coupling, and is very effective in all respects. With careful isolation between the two sets of electrodes operation remains good at very short wavelengths, but in general such combined valves become inferior to two separate stages at wavelengths below about 20 metres. At broadcast wavelengths they form both an efficient and an economical method of frequency-changing, and are found in the great majority of receivers in use to-day. Fig. 114 shows the complete circuit diagram of an octode frequency-changing stage of simplest type.

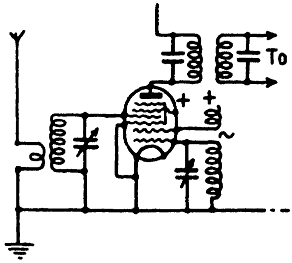


FIG. 114. The Octode.

The pentagrid and heptode are somewhat similarly constructed, the electron stream passing successively through a number of grids grouped in different order in various types. The triode pentode and triode hexode differ fundamentally in that the cathode emission is divided into two streams. One of these feeds the triode elements only, whilst the other feeds

the pentode electrodes separately. The two are isolated and may be treated as two independent valves in one envelope, being coupled externally. In the hexode valve an extra injector grid is added. This is joined to the triode grid internally and serves to couple oscillatory potentials from the latter into the detector electron-stream. In the triode hexode is said to be found the most efficient combination of efficient frequency-changing with isolation between oscillator and detector tuned circuits. It is also more efficient at short waves than most earlier types.

The efficiency of a frequency-changer in transforming signal energy into intermediate-frequency energy is defined by a factor termed 'conversion conductance'. This is strictly equivalent to mutual conductance in an ordinary amplifying valve, no account being taken of the fact that input and output are at a different frequency. The conversion conductance will only be a maximum if the oscillator section is working correctly and is delivering a certain oscillatory potential termed the 'optimum heterodyne voltage' to the detector section. This voltage is usually stated by the valve-makers, and in designing a frequency-changer circuit it is necessary to see that the external oscillator circuits are efficient enough to generate it. At short waves in particular the oscillator may produce insufficient amplitude, which should therefore be measured. Since the conversion conductance is likely to fall off in any case for wavelengths below some 20 metres, it is particularly necessary to maintain optimum heterodyne conditions. In some valves the mutual conductance of the oscillator electrodes is poor compared to a standard triode, and short-wave performance can be improved either by reverting to a separate oscillator or by joining an external triode in parallel with the oscillator section. More intense oscillation is thus produced.

A fault in early frequency-changing circuits was the tendency to generate spurious harmonics or beats. This has also been largely eliminated by careful attention to the characteristics of such special valves as the triode hexode.

The signals will now be at the intermediate frequency, and a coupling suited to the new frequency must follow the first detector. This nearly always takes the form of a loosely coupled transformer, the tuned primary of which completes the anode circuit of that stage, as shown in the diagram, whilst the tuned

secondary passes on the I.F. potentials to the grid of the first I.F. stage. It is important that this first coupling should be selective, and in particular should pass very little energy at frequencies removed from the intermediate. The process of frequency-changing by heterodyne gives rise to a number of components at other new frequencies, some of which may be of considerable amplitude, and it is most desirable to prevent these from entering the I.F. amplifier. Second-channel interference is not the only kind encountered when a signal frequency X is heterodyned by an oscillator frequency Y, although it is the most serious. The heterodyning process can produce an intermediate difference frequency $(X \pm Y)$, but also beats are possible between harmonics of the local oscillator and the incoming signal, or against harmonics of any other signals which may reach the first detector. Yet another range of possible beats may cause trouble, since the rectification of the I.F. itself by a non-linear second detector will give rise to a distorted wave form containing strong harmonics of the intermediate frequency. These may be radiated from the following circuits and may beat with the original oscillator to give audible tones. This effect is most likely to give trouble when the incoming signals are of very high frequency and the selectivity low, as in television reception. It might be imagined that such a wide range of possible beats would make clear superheterodyne operation impossible, but fortunately most of these are of small amplitude, and with reasonably careful design and screening throughout they can be reduced to negligible proportions. On occasions, however, special receiver requirements may bring one or other of them into prominence, and so their existence should be noted.

The I.F. amplifier is both the most characteristic and one of the most important features of the superheterodyne. It is designed on similar lines to the high-frequency amplifiers which have already been described, with the simplification that tuning need not be variable beyond a limited range of fine adjustment. The amplification possible here is limited by noise arising in the preceding stages, being of the order of from 1,000 to 3,000 in an average set. In receivers of moderate sensitivity, particularly if signal-frequency amplification is used, a single I.F. stage having a gain of from 100 to 300 often suffices; whilst highly selective and sensitive receivers will employ two or occasionally three stages. It is only when specially broad frequency response

is required that still larger numbers become necessary, and the stage gain and consequent number of stages are bound up fundamentally with selectivity. It will therefore be convenient to defer the fuller discussion of I.F. amplification until the following chapter.

After adequate amplification the I.F. signal must be finally rectified, being still a modulated radio-frequency carrier differing only in frequency from the original signal. This is done by the 'second detector', which in earlier designs might be of any of the types that have been described, and might operate with quite a low input amplitude. It has been found preferable from all angles, however, to employ diode detection, working at relatively high level, and the diode is now invariably met with in modern receivers. The technique of stable I.F. amplification has so far advanced that no difficulty is experienced in delivering an I.F. signal amplitude of several volts R.M.S. to the second detector. At this level, diode rectification is the most distortionless and simplest type, whilst offering subsidiary advantages to be mentioned shortly. High-level detection reduces the need for low-frequency amplification, thus avoiding noise or hum, which we have seen to be most easily introduced at this point. The low-frequency stage or stages will differ in no respect from those used after the detector of a straight receiver.

Up to the present no mention has been made of sensitivity or volume control. In the simplest receivers, such as the crystal detector, signals will seldom be so strong as to need reduction. A certain degree of control is permissible by detuning in the case of such unselective receivers, the quality of telephony being little affected by this; whilst in the case of the reactive valve detector a wide range of control over the weaker signals is given by the reaction adjustment. Where such simple receivers are followed by low-frequency stages, the volume obtained from any station is readily controlled by a potentiometer, then termed the low-frequency volume control. The signal potentials are applied across the whole of this potentiometer, as from the detector stage V_1 in Fig. 115, and the slider S taps off any required proportion of this potential to the following grid G. V_2 is preferably the first low-frequency stage, as if it be a subsequent one there is a danger of the earlier stages overloading from very strong transmissions. This volume control may very well replace the grid leak of a R.C.-coupled stage, when its total

resistance will be equal to that of the original leak. This simple arrangement is satisfactory for the control of volume from a simple receiver, provided that the strongest signals likely to be met with will not overload the detector itself. Where there is no signal-frequency amplification this will be unlikely.

In the case of sensitive and highly selective receivers, however, more comprehensive methods are necessary. Quite weak signals may be amplified excessively if the receiver be used permanently at maximum efficiency. Reduction of volume by detuning is absolutely prohibited if the receiver be selective. For one reason, it may bring in interference from adjacent signals, which will only be at a minimum when the receiver is accurately tuned to the carrier of the wanted station; and, for a second, it will introduce distortion in telephonic reproduction because of unequal response to the two side bands. Volume must be controlled by the adjustment of actual amplification, this being reduced until signals reach the telephones at the desired strength. Preferably the method of adjustment must be such that it has no effect upon selectivity or quality, a requirement which is not met by adjustment of reaction or by detuning.

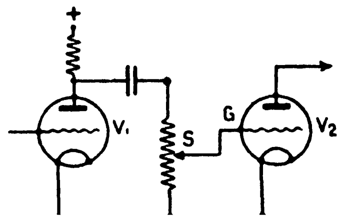


FIG. 115. Volume Control at Low Frequency.

A sharp distinction should be drawn between control of volume and of sensitivity. The fact that in simple cases the two are not easily distinguished has led to much loose talking, the term 'volume control' being used for anything which varies the output of a receiver. Control of low-frequency amplification can correctly be termed volume control, because the amplifier can only magnify that which is rectified by the detector. The true sensitivity of the set to weak signals is not changed, and a signal will not be heard unless it was already strong enough to operate the detector. If sensitivity is defined as the minimum input potential which will produce any measurable signal at the output terminals, then a marked difference exists between a low-frequency volume control and one which controls amplification before the detector. In the case of superheterodynes the second detector is of course implied. The writer therefore prefers to

apply the term 'volume control' to an audio-frequency control situated after the final detector, and to use the expression 'sensitivity control', or 'gain control', as it is often termed to-day, for any device which controls the amplification before detection.

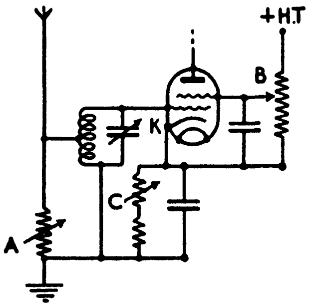


Fig. 116. Sensitivity Control.

A very fundamental method of control would be to attenuate the incoming signals before the first valve. This was done in early equipment by an adjustable 'loose coupler', but since this may affect selectivity or tuning it is not considered convenient to-day. A method now used may employ simply a variable resistance between aerial and earth, or a potentiometer controlling the input to the first grid. This is shown at A in Fig. 116, a composite diagram in which a variety of control systems are shown in a single circuit. They would not of course be combined in practice in a single receiver. Signal input control has one very important advantage, that of definitely preventing overload, however strong the signals from the aerial may be. It has also a fundamental disadvantage when used alone, because, since the receiver will be working at full gain whilst input is reduced, all stations are treated as weak signals, and the signal-to-noise ratio is at its worst. The system is thus ideal for low-gain receivers, in which noise is not a limiting factor, but unsuited to sensitive types, unless it be combined with an amplification control, to which it may be ganged.

The straight receiver using signal-frequency stages comes next for consideration. Here the amplification of these stages, or of one of them, will be controlled. A satisfactory system very general a few years ago was to vary the screen-grid potential of these as shown at B. The magnification given by a screen-grid stage falls off fairly linearly as this potential is reduced from its optimum value, becoming negligible for zero screen-grid bias. Other methods of reducing amplification are occasionally met with. An example is found in a Canadian design, in which the coupling between all the radio-frequency transformers is simultaneously reduced, the method being to increase the separation between primary and secondary. Amplification can be reduced

by a potentiometer at the input of any stage, or by the use of a variable damping resistance across any of the coupling circuits, but such methods may act injuriously upon selectivity. It is also inadvisable to reduce gain by reduction of high-tension voltage, or any similar method which might lead to overload on strong signals. To minimize this possibility the control should take place early in the receiver, or should act simultaneously upon all high-frequency stages.

More recently a special type of valve construction has been developed for both high-frequency pentodes and screen-grid valves, by which mutual conductance becomes dependent upon grid bias. To achieve this the spacing of the control-grid mesh varies along its length from an open formation at one end to a close one at the other, instead of being uniform as formerly employed. Such valves are termed 'variable mu', from the Greek letter used to denote amplification factor. The action is not easily explained in a few words, but becomes evident if it be borne in mind that the 'control factor' of a grid increases with closer mesh, the amplification factor increasing proportionately. Also the bias required to cut off all anode current increases as grid mesh becomes more open. A grid of varying mesh will be entirely operative at low grid bias, giving a control factor proportional to the mean mesh spacing, and the efficiency of the valve will be a maximum. As grid bias is increased, a point will be reached where the closest mesh portion becomes over-biased and ceases to pass the electron stream. The remainder of the grid will still be operative, but since the mean mesh will now be more open the amplification factor will be lower. This process of decreasing amplification factor with increasing grid bias will be progressive, and with correct grid design will be linear, whilst it will not be accompanied by distortion because the bias at any period will be appropriate to the *mean* mesh spacing of the active portion of the grid. At the extreme maximum bias only the most open portion will pass electrons, the remainder being over-biased to cut-off. The amplification will be a minimum, and, owing to the reduced active electrode area, R_a will be high, with a very low resulting mutual conductance. Further increase in bias will eventually cut off all anode current, rendering the valve dead, but before this point is reached distortion and non-linearity occur. Thus it is not practicable to work a variable-mu valve quite down to zero amplification. The

properties of different examples naturally vary, but a typical valve may have a mutual conductance of 3·5 with optimum grid bias of about 3 volts. This will fall linearly to some figure such as 0·1 for a bias of 20 or 30 volts negative.

The variable-mu valve forms an excellent method of sensitivity control with minimum distortion, and is so convenient that it has become the standard method. A variation of amplification by as much as 100 to 1 is possible with reasonable linearity from one stage, although a smaller ratio is to be preferred. Control now takes the form of a grid-bias adjustment, which may comprise a battery and potentiometer, or, in the case of indirectly heated valves, may be a variable cathode resistance, as shown at *C* in Fig. 116. A low value of fixed resistance is inserted in series with this, so that optimum bias remains when the control resistance is completely cut out. A straight receiver employing two high-frequency stages can be effectively controlled by this method, provided the low-frequency volume control is also used in the case of the strongest signals. It may also be necessary in cases where a local station is received at exceptional strength to provide a measure of input control, mainly to prevent overload of the first grid. This often takes the form of a fixed resistance across some part of the input circuit, which is cut in by a 'local-distance switch' when it is desired to receive the local station with good quality. Progressive improvements in variable-mu valve design have made this switch unnecessary in the average modern receiver.

The variable-mu valve has made practicable the important development of automatic volume control, abbreviated to A.V.C. and more correctly termed automatic gain or amplification control. The essential idea underlying this circuit is that an incoming signal shall itself adjust the sensitivity of the receiver, in such a way that a uniform output level results regardless of the signal intensity at the first grid. Clearly, if the signal amplitude can be made to control amplification so that the two are inversely proportional to each other, the respective rates of change being equal, then no variation in output would occur under any conditions. The effect of any increase in the input will be exactly compensated by the reduction of amplification which it produces. This condition could be exactly obtained if the incoming signals were amplified by two similar receivers, the first of which provided audible signals in the usual way,

whilst the second served solely as a control channel. In the latter the amplified signals would be rectified by a diode detector, which we have seen produces not only a reproduction of any modulation but also a steady current proportional to the carrier amplitude. If the sensitivity of this second receiver be unaltered, then the rectified diode current will at all times be proportional to the amplitude of the signal carrier wave at the input of either

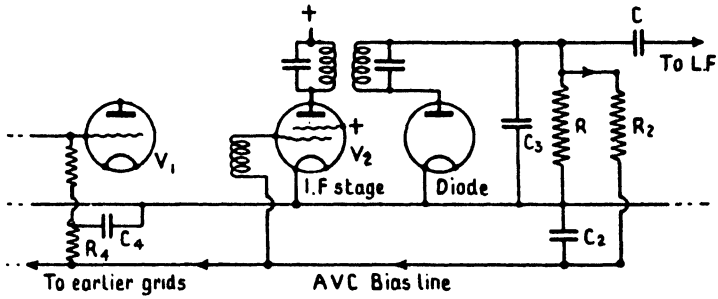


FIG. 117. An A.V.C. System.

receiver. Let the amplification of the first or normal receiver be controlled by the use of variable-mu valves, then the application of a negative control bias to these valves will reduce the final output. This bias can be provided directly from the rectified carrier produced in the second or control receiver flowing through a high resistance, and will thus rise and fall exactly in proportion to the incoming signals. It will then control the amplification of the first receiver also, and if the response of both be in all respects linear, and the rate of control suitably adjusted, the final output from the former will remain constant.

This principle employing a separate control amplifier is the most perfect, and is used in some commercial installations where cost is unimportant. In the average receiver, however, a simpler and slightly less effective circuit gives adequate control. It is necessary that the detector used, or the second detector of a superheterodyne, be a diode. This is one of the main reasons for the wide adoption of this form of detection in modern sets. The products of rectification will appear as potential variations across the load resistance R of Fig. 117, which shows the essentials of an A.V.C. system. These will comprise the rectified modulation potentials, if any, which, being alternating, will pass through the condenser C to the grid of the first

low-frequency amplifier. The current represented by the rectified mean carrier wave will also flow through R, building up across it a direct potential proportional to the strength of the signal. This potential will pass through the high resistance R_2, and will exist across the larger condenser C_2; but the alternating potentials will tend to follow the lower-impedance path through C rather than the high-resistance path through R_2, and any traces which reach C_2 will represent a negligible potential across the low reactance of the larger condenser. As an example of practical values, R may be from 0·25 to 1·0 megohm, $R_2 = 1{\cdot}0$ megohm, $C = 0{\cdot}1$ mfd., C_2 0·1 to 1·0 mfd., and the condenser C_3, which serves merely to by-pass carrier-frequency current across R, from 100 to 500 mmfd. only. Under these conditions the direct potential across C_2 will be substantially free from alternating components at either carrier or modulation frequencies. It may be termed the 'control bias'.

This bias will now be applied to the grids of those preceding stages which consist of variable-mu valves. To do this, the low-potential ends of the respective grid circuits will be connected to the condenser C_2. It may be permissible to make direct connexion, as shown in the case of valve V_2 in Fig. 117, but in other cases, such as that represented by V_1, it is advisable to insert a second resistance and condenser R_4 and C_4 in the manner shown. This combination forms a simple type of low-pass filter, and occurs so frequently in radio-receiving circuits that by now it will be familiar to the reader. Alternating potentials which might exist across the condenser C_2 owing to the operation of the valves V_1 or V_2 must not be allowed to reach other parts of the receiver, where they might give rise to unwanted regeneration or couplings. This may occur if several grid circuits are taken directly to C_2, for whilst the reactance of this is very low at signal frequencies it may still be large enough to act as a common coupling impedance. Hence oscillatory energy from either grid circuit is provided with a low-impedance path to earth through the condenser C_4, which will ensure only a very small proportion passing through the relatively high-resistance path R_4 back to C_2. In this way the potential across the latter reaches each grid, whilst the transfer of oscillatory energy in the reverse direction is checked.

The operation of this type of A.V.C. circuit is easy to visualize. Considering the case of very weak incoming signals, there will

be little potential built up across *R*, and so negligible bias applied to the grids of the variable-mu valves, which may be termed the controlled stages. The amplification of these will then be a maximum. On the arrival of a somewhat stronger signal, appreciable potential will exist across *R* and will apply a corresponding negative bias to the controlled valves, reducing their amplification. This will in turn reduce the incoming signal amplitude reaching the diode, until an equilibrium condition is reached at which this amplitude is slightly greater than that of the very weak signal first considered. A strong signal will produce the same changes in more marked degree, producing a large bias across *R* and reducing the amplification very considerably. Equilibrium will again be reached, but at a slightly greater output from the diode than before. Thus there is a tendency for all signals to be reduced to a common level, but this cannot quite be achieved without the use of a separate control amplifier, since a strong signal must remain slightly stronger at the diode than a weak one if the automatic control bias is to differentiate between the two cases, as clearly it must if the circuit is to operate. If the low-frequency amplification following the diode detector be set to a value giving pleasant reproduction from an average incoming signal, all other signals will be of the same order of volume over a wide range of amplitude at the aerial circuit. This will only apply to very weak signals, however, if there is sufficient amplification in reserve to bring them up to the equilibrium state. Still weaker signals will have no effect on the A.V.C., and its full benefit will thus only be felt when fitted to a receiver of high initial sensitivity. It must be fully realized also that by signal strength or amplitude is meant that of the carrier wave, rather than of the modulation, and that A.V.C. is essentially a method of reducing changes in carrier strength. If therefore one signal is more fully modulated than another, it will be reproduced at greater volume, the control circuits having no effect in changing modulation percentage.

Since transmissions differ in modulation depth, and A.V.C. action will not entirely reduce all signals to a mean level, it is usual to provide a low-frequency volume control to receivers so fitted. This enables the user to adjust the volume of sound to his requirements at all times, and it will not usually need readjustment when tuning from station to station. An R.F. sensitivity control is rendered unnecessary when A.V.C. is employed.

With this fact is bound up rapidity of A.V.C. action. This depends upon the time constant of the resistance-capacity combination R_2C_2, since any change in the signal strength will apply a charging potential to C_2 through the resistance R_2, and will take an appreciable time to complete the charging or discharging process. Actual figures mean little here owing to the very varied magnitude of the amplitude changes concerned, but it is an important factor in design that a large capacity and high resistance will yield a slow-acting control which may take several seconds to reach equilibrium, whilst by reducing both to a minimum the control can be made so quick-acting that low modulation frequencies may be largely 'controlled out'. An intermediate condition will naturally be selected in practice, a good compromise being one that operates as quickly as possible without attenuating the lowest modulation frequencies which the equipment as a whole is intended to reproduce. This will imply a slower action in the design of a high-quality broadcast receiver than in the case of a short-wave or commercial receiver intended only to receive intelligible speech. A suitable time period for the former would be from 0·2 to 0·05 second, whilst the latter might be less than 0·02 second.

The possibility of obtaining a uniform output by A.V.C. methods is increased as the number of controlled stages becomes greater. In any case, the system should be looked upon as a method for scaling down changes of output in relation to those of input; both will still rise and fall together, but the output will do so on a considerably reduced scale. It is not possible to get sufficient control from one stage to provide uniform output from the wide range of signal intensities encountered in practice, and for this reason automatic control is not often fitted to simple straight receivers. With two controlled stages a good measure of control is possible, making A.V.C. satisfactory when applied to more sensitive straight sets and to simple superheterodynes, except in the case of extremes of signal strength. The circuit is most useful in the design of large superheterodyne receivers, in which three or more stages can be controlled—for example, as one R.F. and two I.F. stages, or perhaps one R.F., one I.F., and the first detector. It is a growing practice to design frequency-changers in which the pentode section has a variable-mu characteristic and can be used to increase A.V.C. control in the simpler receivers.

Signals near to the limit of receiver sensitivity will not be improved by A.V.C., and such weak signals will be better received if the control is rendered inoperative until a certain moderate signal level is reached. This is done by the use of a separate diode detector to provide the control bias, provided with a bias sufficient to prevent its passing current until the signal amplitude exceeds a predetermined value. Such systems are termed 'delayed A.V.C.' and are characterized by a control which is inoperative when signals are weak but becomes fully operative if they reach sufficient amplitude to make A.V.C. desirable. 'Amplified A.V.C.' may also be found in receiver specifications. This merely implies a system in which the control bias built up across R is amplified by an additional valve before being applied to the controlled stages, thus improving the effectiveness of the whole system, particularly when a limited number of stages exist.

An advantage of A.V.C. is to minimize fading. This it can do very effectively when signals are of good working strength at the aerial, preventing any objectionable peaks of output. It should be borne in mind, however, that automatic control cannot reduce background noises, but may even make these more evident by ensuring that the receiver works at high sensitivity during periods of fading. Similarly, it cannot reduce interference of any type, but tends to make the receiver noisy during adjustment, since sensitivity automatically becomes high in the absence of a signal, bringing in noise that might otherwise be overlooked. A second use of A.V.C. is that of apparently flattening the tuning of selective receivers and reducing the importance of the manual sensitivity control, both of which assist the unskilled operator.

There is a variety of other automatic-control circuits which can only be briefly touched upon in a work of this scope. Many of these are still in a state of development. One of them has a number of names, of which 'inter-station noise-suppression' is self-explanatory. Here the control bias, derived as before from a diode detector, is used to operate a relay, either mechanical or electronic, which normally cuts out some part of the receiver such as the speaker or low-frequency amplifier. The circuit is then dead in the absence of an incoming signal, and tuning is carried out in silence until a station is received. If of sufficient amplitude to overcome local background, provide a control bias,

and thus work the relay, the receiver is then put into full operation by the signal itself, and remains so as long as a sufficient working signal continues to come in. It will be necessary to introduce a degree of time delay into this circuit, so that the control will not function too rapidly, and will not respond to atmospherics or very brief fading.

Inverted in action and working at much higher rate, a similar circuit constitutes the latest type of 'noise-suppressor', able to reduce atmospherics and noises due to electrical machinery provided that these are stronger than the incoming signals and of very brief duration. The crackling produced by the ignition system of motor-cars is typical of this variety of interference and is most troublesome on ultra-short wavelengths. Adjusted for this purpose, the control circuit is delayed by a bias preventing operation until the amplitude of an interfering impulse reaches a high level, when it is so quick-acting that the relay closes before the disturbance has time to reach the receiver output. A very brief 'hole' is thus punched in the programme heard from the loud speaker, during which the interfering impulse passes unheard. If such interruptions do not exceed about 1/50 second in duration, they are imperceptible to the human ear, and the interfering impulse will be erased from the programme.

Automatic control of tuning is also possible, and is receiving much research and attention to-day. For some time it has been usual to provide a tuning indicator, whereby the operator has visual indication of correct resonance. This may consist of a simple anode-current meter inserted at some point such as the anode circuit of an intermediate-frequency stage, where current will vary somewhat as a station is tuned in; or it may operate from the control-bias circuit. Any current change which reaches a maximum or minimum at exact resonance will work a variety of ingenious indicators designed to appeal to the listening public. The simple milliammeter movement may control a coloured shutter or partly reveal the light from a small lamp, causing an illuminated window to vary as stations are tuned. A special neon lamp has been designed in which the glow will spread along one electrode in proportion to signal amplitude, whilst several small cathode-ray tubes have been pressed into service as attractive indicators, the 'electric eye' being a recent example.

Of more importance is true 'automatic tuning compensation', which should not be confused with the many systems of remote tuning control. This resembles A.V.C., in that the incoming signal itself is made to perform the finer adjustments of tuning after the receiver has been set approximately to resonance by hand. It is also somewhat similar in principle. Two diodes in addition to the main detector are usually employed and are fed by two circuits resonant each a few kilocycles above and below the intermediate frequency respectively. Thus a differential control bias is obtained, that from one diode increasing at the expense of the other if the oscillator be so set that an excessive intermediate frequency is generated, whilst the reverse occurs if the frequency is too low. At correct tuning, when the intermediate frequency is exactly that of the intermediate-frequency amplifier, both diodes are equally affected; no differential bias therefore exists, and the compensating circuits will not act. Assuming an initial detuning of the oscillator, however, there will be a predominance of bias of one polarity or the other, and this actuates a control valve having the ability to change the oscillator frequency until correct resonance is reached. The control then ceases.

There are numerous methods of effecting the latter portion of the process, namely the changing of oscillator frequency by the action of a control bias. Mechanical methods are very effective. Thus, for example, the control bias is converted to a current variation by application to the grid of a power valve. In the anode circuit of this is a meter movement, in which the pointer carries one light aluminium plate of a small variable condenser connected in parallel with the oscillator tuning condenser. As the differential bias acts, this pointer will move one way or the other, raising or lowering the oscillator frequency to compensate for initial misadjustment. Electronic methods are equally possible, most of which rely upon the Miller effect. This states that the effective grid-cathode impedance of a valve varies approximately as $1/\mu$ times the anode impedance. Therefore, if the amplification factor be varied by alterations to grid control bias, the effective impedance formed by the valve across some part of the oscillator resonant circuit will be changed. Automatic tuning is an important field, since it overcomes the difficulties of tuning very short waves or intensely selective receivers by hand, and the still more important difficulty of maintaining them in

exact resonance should the transmitter or oscillator drift in frequency. It has not yet become a standardized technique, however, being in the early stages of research. It is impossible here to indicate the many methods of attack, or the problems which are as yet unsolved. Not the least of the latter lies in the fact that should the incoming signal fade temporarily to zero the control circuits may lose their influence, and a neighbouring powerful station may automatically tune the receiver to itself. We must content ourselves with indicating in bare outline the lines on which this important new development is progressing, and suggesting that it is an interesting subject for study.

EXAMINATION QUESTIONS

1. Why are coupled circuits used in radio-receivers? What is the advantage of variable coupling in a receiver? Indicate with diagrams two methods of providing variable coupling in a receiver. What is meant by the term 'coefficient of coupling'?
City and Guilds of London Institute. Preliminary Exam. 1937.

2. Why is retroaction sometimes used in receivers? Indicate two methods of applying retroaction to receiving circuits. What are the disadvantages, if any, attending its use?
C. and G. of L. I. Preliminary Exam. 1936.

3. Explain, with diagram, how automatic volume control is obtained in a receiver. *Institute of Wireless Technology. May 1936.*

4. Compare the advantages of radio-frequency and low-frequency amplification in a radio-receiver of the straight type.

5. Explain the principle of super-regeneration. What are the principal characteristics of this circuit? Indicate one method by which it can be carried out.

6. Explain the use of an 'anode tap' in a valve amplifier. If a split condenser is used for this purpose, what is the relation between the ratio of capacities and the ratio of the impedances involved? (See also Chapter VI.) *A. M. I. W. T. June 1937.*

7. Explain fully the theory of the heterodyne reception of continuous-wave wireless-telegraph signals, and describe, with a circuit diagram, how it is carried out in practice. *Grad. I. E. E. 1935.*

8. Describe with a diagram the principles of operation of a super-regenerative type of receiver. What are the advantages of this type of receiver, and what is the range of wavelength for which it is best suited? *C. and G. of L. I. Intermediate 1935.*

9. Explain the relationship between the dynamic resistance, the selectivity, and the Q of a tuned circuit.

10. What is a wave trap? Sketch the connexion to an aerial circuit of a series-acceptor and a parallel-rejector type of wave trap. For what purposes are they suitable?

11. What is 'neutralization'? How can it be carried out? For what purposes was neutralization introduced, and for what purposes is it still valuable to-day?

12. Compare three methods for controlling the sensitivity (volume) of a radio-receiver, mentioning the class of set to which each is suited and their relative merits.

13. Explain the principle of superheterodyne reception.

14. How is the frequency of an incoming signal changed to a fixed intermediate frequency in the superheterodyne? Describe three modern types of 'frequency-changer'.

15. What is meant by the terms (*a*) preselection, (*b*) second detector, (*c*) second-channel interference, in superheterodyne reception? Explain how (*c*) occurs and its relation to (*a*).

CHAPTER XI

SELECTIVITY IN RADIO-RECEPTION

SELECTIVITY, or the ability to separate one radio-signal from others on closely adjacent wavelengths, has been mentioned on several occasions in the preceding chapters. With the steady increase in the number of active radio-stations during recent years, all of which must be accommodated within a frequency spectrum limited by considerations of propagation, it has been found essential to reduce the frequency separation between stations to the minimum. In spite of careful regulation by international agreement, whereby nearly all transmissions are now accurately maintained upon selected frequencies suited to their needs and those of their neighbours, considerable interference is bound to occur. This is probably on the increase, and it may not be an exaggeration to say that the related problems of selectivity and interference between transmissions have become the most important individual factors in radio communication. They are also the chief factors which demand care in the design in both transmitters and receivers, and for these reasons it has been thought best to devote a separate chapter to their discussion.

In the beginnings of radio, selectivity was of little account. Transmissions were few, and their range limited by very poor receiving sensitivity, so that the problem was mainly that of hearing something of the wanted signals. Often there would be no other transmitter within range, or at most a few on widely different frequencies. Tuning, or syntony as it was then called, was introduced by Lodge as an important step in the improving of sensitivity. A tuned transmitter was found to radiate much more energy and in a more effective manner than the earlier untuned systems, whilst tuning the receiving aerial brought in signals from hitherto unexpected distances. The ability to separate transmissions which it brought with it was found convenient as the range and number of these increased, but was for a time quite secondary. The simple coherer or crystal set, for example, with which our chapter on receivers opened, would be unable to separate more than some half-dozen spark transmissions within the scope of the present broadcast waveband,

provided of course that these were all of considerable strength. Modern undamped or telephony transmissions would be somewhat more easily separated, since they do not spread over so wide a frequency band. Early spark transmitters possessed a distinctive musical note, and although several would often be heard in the telephones at one time, the skilled operator developed a surprising ability to read one and disregard the others. Continuous wave or telephony transmissions do not lend themselves so well to this treatment.

Sideband Theory. With the development of the valve receiver sensitivity ceased to be the limiting factor. The range of transmission soon became worldwide, and interference between the numerous signals now made audible for the first time became serious. Selective circuits were developed as quickly as possible, and for their own sake. At first it was believed that no limit should exist, and that a sufficiently selective receiver would separate any pair of transmissions not actually on an identical frequency, but experiment showed that this is not the case. The explanation lies in the Sideband Theory explained briefly in Chapter IX. There we saw that if a carrier wave of frequency f be modulated by a frequency p, then two new frequencies of $f+p$ and $f-p$ come into existence. A simple mathematical analysis shows the exact equivalence of a carrier wave of constant frequency f modulated by a sinusoidal wave p on the one hand; and the combination of a uniform carrier f with the two 'sidebands' $f+p$ and $f-p$ on the other. It cannot be too strongly stressed that the two are in fact different statements of the same physical fact. Both are strictly true descriptions of the modulated wave, always equally applicable, and should lead to the same result if employed in any valid argument. No importance therefore attaches to theories in which the one is regarded as true at the expense of the other, such as statements that under certain conditions 'there are no sidebands'. We are dealing with an identity. Whenever modulation by the frequency p occurs, energy must exist at the upper and lower sideband frequencies. We may not choose to make use of these, and may even suppress one or the other by dissipating its energy in filter circuits; but there can be no question of their existence as real and measurable oscillations.

The existence of sidebands is shown by the following simple trigonometrical analysis. The equation of a sinusoidal wave

can be written $E = E_0 \sin \omega t$, where E_0 is the initial amplitude, E that after a time t, and $\omega = 2\pi f$. If this be modulated by an oscillation of frequency p to a depth given by a modulation factor M, the equation becomes

$$E = E_0(1+M \sin \omega_2 t)\sin \omega t,$$

where $\omega_2 = 2\pi p$. Multiplying out we obtain

$$E = E_0 \sin \omega t + ME_0 \sin \omega t \sin \omega_2 t.$$

We can expand the latter term by a well-known transformation, obtaining

$$E = E_0 \sin \omega t + \tfrac{1}{2}ME_0 \cos(\omega - \omega_2)t - \tfrac{1}{2}ME_0 \cos(\omega + \omega_2)t.$$

This expression is the sum of three separate waves, one the original or carrier wave $E_0 \sin \omega t$, and the others differing in frequency by the modulation frequency ω_2, and each of $\frac{1}{2}ME_0$ times the amplitude of the original wave. These are the sideband frequencies, previously called $f+p$ and $f-p$. Further, when $M = 1$ for complete modulation, we see that the amplitude of each sideband becomes half that of the carrier wave.

The existence of these components of the modulated wave can be simply shown in the laboratory. An oscillator at any convenient frequency f is set up, and is modulated by a frequency p which for convenience is about 10 per cent. of f. Then if a wavemeter be brought near to the oscillator, radiation at the frequency f can be detected, and on retuning the meter to the frequencies $f+p$ or $f-p$ a reading will also occur, showing that actual radiation at the sideband frequencies is taking place. It will be found that for 100 per cent. modulation, the energy of each sideband will be half that of the carrier, the total energy being divided in the ratio of one-half at the carrier frequency and one-quarter at each of the sidebands.

It may seem wrong to talk of 'bands' when only a single side frequency of $f+p$ or $f-p$ is referred to, but this is done to indicate that the theory applies equally to modulation by any number of sinusoidal frequencies simultaneously. Each of these modulating waves will produce its own pair of 'sidebands', so that altogether there will be a number of these occupying a 'band' of frequencies on each side of the carrier, up to a distance represented by the highest individual modulation frequency. Now it has been pointed out that complex modulation such as telephony can be analysed into the sum of a large number of pure

sine waves, and will therefore give rise to a family of sideband frequencies each of which obeys the laws that have been stated. Perhaps the case of an orchestra will help to visualize this. There are numerous instrumentalists each producing at a particular instant one signal note, which may approximate to a pure tone, or be accompanied by a series of overtones. A group of sine waves will thus represent each instrument, whilst the larger group formed by adding all these will represent the sound of the full orchestra, and can be transmitted as a signal complex modulation giving rise to sidebands stretching from the carrier frequency f up to the highest frequency present in the orchestra.

A telephony transmission will thus occupy a definite band in the radio-frequency spectrum, equal to twice the width of one sideband. This is a fundamental fact which must be recognized, and of such importance that it has been explained here once again at the risk of repetition. Were it possible to design a radio-receiver so selective that only the carrier frequency and no other would be heard, telephony would be impossible, since the sidebands would not be received. These contain *all* the energy of modulation, and so the carrier wave would be 'demodulated', or converted back to its original unmodulated state. As has been pointed out in Chapter IX, telephony can be radiated as sidebands only, the carrier necessary to heterodyne these and make them audible being added at the receiver. The fact that this can be and is done furnishes indisputable proof that sidebands are a reality, and not of purely theoretical importance.

It will be noticed that a low modulating frequency will result in a sideband near to the carrier wave. In any case of course when the carrier is a high radio-frequency, such as 1,000 Kc., corresponding to a wavelength of 300 metres, the sidebands of ordinary telephony will only represent some 1 per cent. or less of the carrier frequency. Within this narrow band, however, a low modulating note of, say, 100 cycles per second will only differ from the carrier by this small amount, the sidebands being at 1,000,100 and 999,900 cycles respectively. On the other hand a high note of, say, 10,000 cycles will produce sidebands at 1,010,000 and 990,000 cycles, considerably more remote from the carrier frequency. This is a general principle, that high modulation frequencies, or notes of high pitch, are transmitted by radio as sidebands relatively remote from the carrier, whilst low pitched notes will be transmitted by sidebands

close to the carrier. It now becomes evident that a very sharply tuned or selective receiver may modify these sidebands unequally. Look at the resonance curve of Fig. 25, for a lightly damped long-wave circuit. The carrier frequency is 2,350, rather too low for the present argument. Assume it to be 23,500 cycles, the ordinates being multiplied by ten. Then the sidebands of a 500-cycle modulation note would lie at 24,000 and 23,000 cycles respectively. The response to these of the flat (14-ohm) circuit would be little different from that to the carrier itself, but the sharp (4-ohm) curve is about three times down in amplitude at 500 cycles from resonance, and so would only receive the sidebands at one third their original strength. A 50-cycle sideband would, however, be little reduced, while one of 5,000 cycles would be outside the diagram, and reduced to a very small fraction indeed. Thus it is evident that a receiver as selective as this would not reproduce high modulation frequencies adequately. This is the explanation of the uselessness of low carrier frequencies (long wavelengths) for high quality telephony, for which they are never used.

Knowing now that this defect, termed loosely 'sideband cutting', is to be guarded against, it will be desirable to form an opinion of its seriousness. What are the frequencies involved in practical telephony modulation ? Clearly there is seldom any need for the transmission of sounds too high in frequency for the human ear to detect, even if microphones or loud speakers would reproduce these. The ear of a young man can detect up to from 15,000 to perhaps 30,000 cycles per second, but this range falls off later in life, or if hearing is slightly below normal. A good average figure is 10,000 to 12,000 cycles, which is the highest musical frequency that the trained ear demands for virtually perfect reproduction. Many people find music acceptable if reproduction extends up to 5,000 or 6,000 cycles only, and although noises and speech lose something of their natural tone under those conditions, it can be regarded as commercially satisfactory when no better is available. Intelligible speech, adequate for commercial telephone circuits, need not extend much above 3,000 cycles, and this is about the least usually tolerable.

Telegraphic signalling also involves sideband effects, since every time the carrier wave is interrupted or varied to form a 'dot' in the morse code, modulation takes place. High-speed

mechanical sending at several hundred words a minute may involve sideband frequencies extending a few hundred cycles from the carrier, and so the case is similar to telephony of very low quality. As a rule telegraphic sidebands are a good deal narrower than for telephony, and so it is sufficient to discuss the latter, remembering that the case of telegraphy will be similar but less exacting.

The position therefore is that for telephony sidebands up to some 10,000 cycles on each side of the carrier are desirable, whilst half that range is generally acceptable. Commercial speech and high-speed telegraphy can be carried out with a 3,000-cycle or even narrower sideband. For true reproduction it is necessary that this range be received without loss, and the selectivity curve of a receiver should therefore be practically horizontal over the width of the two sidebands. The 'band width' of an amplifier or receiver is defined either as the frequency band which it will amplify without loss, or as twice the highest frequency contained in the sidebands of a modulated signal passing through it without distortion.

Sharply Peaked Circuits. Now a single resonant circuit will always possess a resonance curve of the shape shown in Fig. 25, characterized by a more or less sharp peak and long 'skirts' which never fall theoretically to zero since the curve is a hyperbola. The sharpness of the curve depends upon the degree of total damping, and provided practical considerations will allow this to be reduced sufficiently, the curve can be sharpened almost indefinitely. The most potent device which will do this is reaction, which has been shown to reduce the effective damping of a circuit to zero, or make it 'negative' with the result of self-oscillation. In any case, however, the general shape of this curve is far from ideal if it be relied upon to provide the selectivity of a receiver. When damping is large, selectivity is clearly poor, and if response is to be nearly level over a band width of even 10 Kc., stations 50 Kc. or more from resonance will still cause serious interference. Suppose the curve be sharpened through reduced damping, then the attenuation of the higher sidebands will become increasingly serious, and telephony will suffer. Moreover, the beneficial effects of low damping have little effect upon the 'skirts' of the curve, and response to stations many kilocycles from resonance remains appreciable. A reduction in response by 100 times at 50 Kc.

from resonance would be considerable for this kind of circuit. It may easily occur that a local station is transmitting at that point, having 10,000 times the field strength of the distant station to which the circuit is tuned. As a result, the local station will be heard at 100 times the strength of that wanted, and will completely 'wipe out' the latter. This kind of trouble is typical of the more simple straight receivers, and of early types relying on one or two tuned circuits only, with or without reaction.

We see that a compromise exists. A sharp resonance curve is desirable to reduce interfering signals, but when sharp enough to do this effectively it will also reduce sideband response, with the result that high modulation tones are weakened or almost lost. A musical programme will become low-toned, or 'woolly', and speech will lose much of its character. Serious distortion has been introduced by the selective circuits, and reproduction will not be satisfactory unless this be compensated in the low-frequency stages.

A better condition exists if several similar resonant circuits are employed in 'cascade', such as in the form of high-frequency couplings between several amplifying stages. To obtain the resonance curve of identical circuits so used, the co-ordinates of the resonance curve of one of them are raised to the power n; being squared for two circuits, raised to the third power for three, and so on. A graphical construction or a little thought will show that the new response curve remains somewhat similar in outline, but differs in proportions. The peak of the curve is somewhat flattened, whilst the sides become relatively steeper, and the 'skirts' lowered. The curve is therefore better able to separate signals. Fig. 118 illustrates typical curves approximately to scale, showing the band width of 10 Kc. which must be evenly reproduced for moderate quality telephony. Curve A might be that of a single circuit of average damping, for purposes of comparison. This will become more like curve B when reaction is applied. The curve given by several curves such as A in cascade will resemble C. Here the peak is wider and sideband attenuation considerably less, resulting in better reproduction of the higher modulation frequencies, combined with less response to adjacent signals.

Band Width. The next aspect of selectivity which influences design is the realization that the band width required to repro-

duce any given modulation is the same, irrespective of carrier frequency or wavelength; whilst the attenuation given by a resonant circuit of specified damping is dependent upon frequency, falling proportionately as the operating frequency is increased. As an illustration, curve *A* or *B* will retain much the same shape and relative proportions if the resonant frequency be changed within wide limits, by the selection of different values of inductance and capacity. Provided the 'goodness' of the circuit and the *L* to *C* ratio remain the same, it is possible to multiply

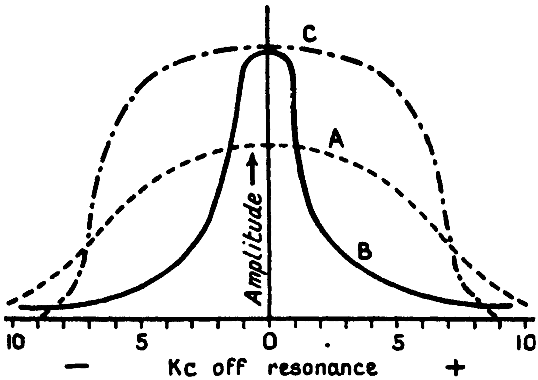


FIG. 118. Typical Resonance Curves.

the frequency scale by any factor, and still retain the true curve of a typical circuit. This of course implies that the ability to separate transmissions will vary enormously at different wavelengths or frequencies. Suppose curve *A* represents the selectivity of a simple receiver. At a carrier frequency of 1,000 Kc. the response will be seen to be about ten times down for a detuning of 10 Kc., or 1 per cent. This will attenuate a 10,000-cycle modulated note five times, or reduce an interfering station 1 per cent. different in frequency by the same amount, approximately. Now consider an equivalent circuit at a carrier frequency of only 100 Kc., 3,000 metres wavelength. One per cent. is now only 1 Kc., and so a 1,000-cycle note is reduced five times, a 10,000-cycle note being reduced perhaps 50-fold or more, resulting in far poorer reproduction. But an interfering carrier at 1 per cent. from resonance is now reduced 50 or more times in relative intensity, and selectivity is thus improved. Alternatively, at 30 metres wavelength where the carrier frequency is 10,000 Kc., 1 per cent. represents 100 Kc. Practically

no loss of the higher modulation notes takes place, but also the circuit is unable to discriminate against an interfering carrier at 10 Kc. by more than a few per cent., the effective selectivity being poor.

The selectivity of a receiver may be defined in several ways. If it is used to imply the proportional response to carrier waves a specified percentage from the resonant frequency, then this 'percentage selectivity' will be the same for given circuits at any incoming wavelength, apart from fairly small variations due to practical details. But we have seen that a band of, say, 10 Kc. is necessary to transmit moderately good telephony at any carrier frequency. It is therefore logical and in accordance with common sense that transmitting station frequencies should be allocated on a basis of equal actual frequency separation. This is in fact done by all governing bodies to-day, carrier frequencies being chosen with regard to the sideband width required for the type of modulation to be used, and with little regard to receiver selectivity. It is assumed that transmission in the case of broadcasting and most other services will employ both sidebands, and that these must not 'overlap' those of the next station in frequency. On this basis European broadcast stations are separated by 9 Kc., which allows a 4,500 maximum modulation frequency without interference; whilst in America the band width selected is 10 Kc., perhaps a more logical figure. These separations are chosen from reasons of economy rather than quality, there being so many stations requiring frequency allocation to-day that it has become necessary to pack them together within the available spectrum at the closest frequency spacing that will allow a reasonable modulation width. Really good quality reproduction of music would thus seem impossible under present broadcast conditions when listening to a distant station, and without interference from powerful neighbours. This is unfortunately often the case, but it must be borne in mind that broadcast stations are intended mainly for reception within the local service area, within which their field strength will greatly exceed that from any other station on an adjacent frequency. It is then possible to extend the width of modulation to perhaps 10,000 cycles, employing lower receiver selectivity, and thus obtain altogether better reproduction.

In the interests of interference, the modulation of any broadcast transmitter should be limited to a maximum frequency of

5,000 cycles, for a carrier separation of 10 Kc. There will then be no 'overlapping' with the sidebands of an adjacent transmission, and the station is said to be operating upon a 'clear channel'. Unfortunately most stations modulate up to much higher limits, and thus cause some interference with adjacent transmissions, since their higher sideband frequencies heterodyne with the other carrier and become audible as modulation upon the latter. This effect is not so serious as it might seem, because the energy contained in the modulation components of most programmes above 5,000 cycles is small, and so the interference is not generally intense. It is therefore neglected in most discussions such as the present, and transmitters are assumed not to radiate sidebands at more than 5 kc. from the carrier.

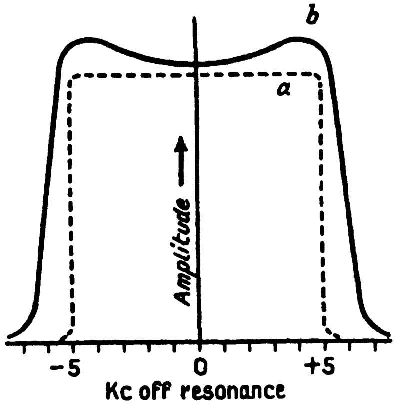


FIG. 119. Band-pass Curve.

The idea of the Clear Channel is an important one, much used in communication engineering. The 'channel' means that band of frequencies occupied by the transmission, including both sidebands if these are radiated; and it is 'clear' if no other transmission exists within that distance on either side of the carrier. We shall regard 10 Kc. as a normal channel throughout the following paragraphs, bearing mind that narrower channels are often sufficient for commercial services. The case of broadcasting has been quoted because it will be familiar to most readers, but identical principles apply to all modulated transmissions.

It has now been made clear that selectivity cannot be increased indefinitely, but is limited by fundamental physical considerations to a band width of one channel, some 10 Kc. in the case of broadcasting. This makes it possible to define selectivity as the ability of a receiver to separate two adjacent channels, a 100 per cent. selective receiver being one which will receive all the modulation of one channel with no interference from the next. Fig. 119 shows by the dotted outline the type of 'resonance' or response curve which would be able to do this. It has long been the aim of designers to obtain such a response,

but so far it can only be approximately attained. It is described as a 'perfect band-pass' response.

The performance possible from the 'straight' receiver, as opposed to the superheterodyne, can now be reviewed from the point of view of selectivity defined in this manner. Early receivers employed one or often two loose-coupled circuits at the incoming carrier frequency only, assisted later after the invention of the triode valve by intensive use of reaction. It has been shown that the selectivity of simple resonant circuits decreases with frequency, and this is a characteristic of straight receivers. Very fair separation may be possible at the long wavelengths, once exclusively used, provided that loss of the higher modulation frequencies is tolerated; but at short wavelengths where the carrier frequency is very high, selectivity becomes poor. With the help of reaction weak signals may be amplified and well separated in the absence of any considerably stronger ones on adjacent frequencies, but if these be present they will spread over the dial and obliterate those wanted. The wave traps and rejector circuits explained in Chapter X were evolved to get over this difficulty. When interference is due to one or two local stations only, these traps can be tuned to them and will result in such reduced input from the aerial at those particular frequencies that the receiver may be able to deal with them. They are also useful to keep out interference from transmitters which may be in the same station buildings, and taking part in the same service as the receiver in question. Whilst still useful for such work, wave traps do not increase the selectivity of the receiver as a whole, and have largely given way to methods which do this.

The next step in receiver evolution was taken by the use of many tuned circuits in cascade, receivers being built with as many as ten of these, individually adjusted by hand. The resulting curve approaches more and more to the ideal shape, and selectivity may become sufficient to separate adjacent channels very adequately at the longer wavelengths. The complexity of such methods has limited their use in the straight receiver, however, although they may be found in the intermediate frequency amplifier of the superheterodyne where they can be adjusted once and for all, whilst the difficulties of ganging so many circuits at short wavelengths makes the system unsuitable.

Band-pass Filters. An alternative arrangement is the Band-pass Filter, the modern name given to a pair of loosely coupled circuits, of the type illustrated in Fig. 29 (*a*) or 30. At suitable values of coupling the frequency difference between the two peaks can be made slightly less than the desired channel, and as shown by *b* in Fig. 119 a better approach to the ideal dotted curve of *a* in Fig. 119 results. Not only is the 'peak' more nearly flat, but the sides of the band-pass curve are steeper and the 'skirts' reduced, assisting selectivity in all respects. Theoretically the band-pass filter seems the best arrangement, and is a characteristic of recent receiver design. Provided a sufficient number of circuits, coupled in pairs in this way, are used, it is possible to approach the ideal rectangular shape sufficiently for most practical purposes. This is possible in straight receivers at the lower carrier frequencies, where four circuits are common, but becomes increasingly difficult at higher frequencies. Difficulty also exists in maintaining the band-pass curve of the correct shape and proportions when a receiver is to be tuned over any considerable wave-band by ganged condensers. The circuits may get slightly out of 'line', resulting in incorrect peak separation. Also the coefficient of coupling must be maintained in correct relationship to the L to C ratio, difficult to attain over any wide range of variation. As a result of these changes, the full benefit of a band-pass curve is seldom realized at all wavelengths to which a straight receiver may be tuned.

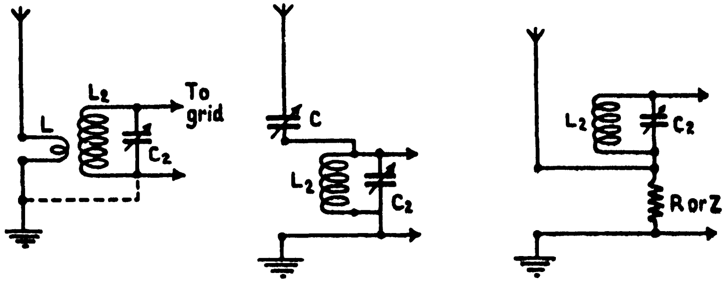
The increased amplification of modern valves makes it unnecessary to couple these in the most efficient manner, and coupling may often be loosened in the interests of selectivity. Thus a step-up transformer ratio may be used, or the anode of a valve tapped across part only of the tuned anode circuit, resulting in either case in less damping and consequently a more sharply resonant circuit. This is a characteristic of selective design, the damping across all resonant circuits being reduced to a minimum, both by loose coupling to the load and the use of low-loss inductances and condensers. The aerial will behave as a considerable load and flatten the resonance curve of any circuit to which it is coupled. At this point also selectivity can be improved. In early receiver design it was usual to either tune the aerial itself, or to couple it very tightly to the first circuit, in the interests of maximum sensitivity.

With almost unlimited amplification available within the receiver, however, this is only now necessary when the weakest signals slightly above local noise level are to be utilized. At other times the aerial will be very loosely coupled, not only to assist selectivity, but to reduce cross-modulation and to make ganging practicable. It has only become possible to gang several circuits accurately with the increasing use of loose coupling between each circuit and the external load. The load impedance represented by an aerial or a valve anode circuit, for example, will be quite different, and will vary differently with frequency, thus unequally affecting the resonant frequency of any given coil and variable condenser to which they are connected. It is most difficult to allow for these variations, and in the case of aerials which may differ widely from each other, it is almost impossible. If, however, coupling be so weakened that the effect of each load becomes small, the differences between them may be neglected, and the circuits will tune similarly.

Aerial Coupling. The original method of aerial tuning at long wavelengths consisted in the addition of inductance to the aerial, and if necessary a series or parallel variable condenser, as in Fig. 24 of Chapter IV. This is only possible if the natural wavelength of the original aerial be not much greater than that to be received, but at short wavelengths the aerial is often too long to make direct tuning possible. It will be shown in Chapter XIV that short-wave aerials are most efficient when they are in themselves resonant to the desired wavelength, being coupled to the receiver or transmitter by a 'transmission line' or 'feeder' which plays no part in the radiating system. The alternative for general reception of a wide range of wavelengths is that of an untuned aerial wire, which may be termed an 'aperiodic aerial'. This would only be directly tuned by added inductance for very long-wave working; seldom when the wavelength lies below 2,000 metres.

The aerial may be loosely coupled to the first resonant circuit by inductive, capacitive, or even resistive methods, illustrated in Fig. 120 at *a*, *b*, and *c* respectively. The coil L in the first case will consist of a few turns only, being resonant at a higher frequency than any to be received, and being tightly coupled to L_2. The actual coupling is weak, however, because the inductance of L is small. In the second case also the coupling condenser C is of low capacity, perhaps between 2 and 50 mmfd.

according to frequency. It may be variable with advantage thus providing an adjustable coupling and perhaps a means of 'trimming' the adjustment of L_2C_2. Both these methods tend to operate best at low wavelengths, the reactance of a small condenser clearly falling with increased frequency. The coupling thus increases with frequency, and when this is undesirable, the resistive arrangement shown in Fig. 120 *c* can be used. Here coupling would be constant if the dynamic resistance of L_2C_2 did not fall, but since this is inevitable for any large reduction



(*a*) (*b*) (*c*)

FIG. 120. Methods of Aerial Coupling.

of frequency, the circuit is little used. It is more important if the resistance be regarded as a mixed impedance Z. If this be an inductance, the circuit resembles *a*, but by using a resonant circuit in place of Z it is possible to make the coupling peak at a particular frequency.

The general tendency in straight receivers is for sensitivity to rise with frequency, because if tuning is by variable condensers the L to C ratio rises as their capacity is reduced. In general it will be better if the overall sensitivity can be uniform with frequency, and this will be approached if the aerial coupling be designed to fall off with frequency. In one method that has been used to achieve this the aerial is taken to a rather large inductance L_3, which in conjunction with its own self-capacity and that of the average aerial will resonate to a wavelength somewhat greater than the longest to be received. The curve of the aerial circuit will be rather flat, damping being added by resistance if necessary to make sure of this, in which case maximum coupling will occur near resonance, and will slowly but progressively fall as the wavelength is reduced below that of L_3.

Apart from the use of low damped circuits, straight receiver design is characterized by the use of band-pass filters whenever practicable. A very popular type of receiver employs one signal frequency stage, a regenerative detector, and low frequency stages as required. In this case a band-pass filter may precede the first valve, whilst a single circuit coupling will be used between this and the detector, rendered sharply peaked when desired by reaction. It is not readily possible to use a second band-pass coupling here, because the application of reaction to one half of this upsets the curve shape, and makes the filter little better than a single circuit. If two signal frequency stages be used, a second band-pass filter will be found between these. In this case reaction may be dispensed with, when a third filter becomes possible before the detector, making a total of six ganged circuits. It is seldom worth while to employ more than this, particularly as the superheterodyne is a better method of reaching an equal or better degree of selectivity, and for this reason receivers employing more than three or four ganged circuits are becoming exceptional. Although the selectivity of these is not extremely great unless they be somewhat elaborately designed, it must not be thought that they are of little use. Their very simplicity is of course in their favour, particularly when either cost or bulk and weight are major considerations. They are amply sensitive for most uses, and have a very good signal-to-noise ratio.

Demodulation. The intensive use of reaction in separating one weak signal from another introduces a new and valuable effect, termed 'demodulation'. This is responsible for much of the efficiency which simple straight receivers certainly possess. It has been shown that the effect of applying reaction until the total circuit damping is very nearly zero is to sharpen the resonance curve until a very pronounced peak exists. This means that the wanted carrier will be enormously more amplified than another a few kilocycles off resonance, and although the latter may be strong enough to be heard at objectionable strength, the ratio of wanted to unwanted signal amplitude will be very high. Now it so happens that when a strong carrier is applied to a detector, the working point of this will be determined almost entirely by that large carrier, and will naturally differ from the no-signal condition. It can be shown that this new condition is very unsuitable for the rectification

of the weaker signal arriving at the same time, and that the latter will behave just as if it were a very weak modulation upon the original strong carrier. It will represent a very small percentage modulation indeed and, provided that the ratio of the two amplitudes be large enough, will become virtually inaudible. An illustration of this effect occurs if a near-by transmitter be started up whilst a moderately strong signal is being received on a simple set. The signals, previously of good strength, suddenly become inaudible, and only those from the local transmitter can be heard. This effect is termed 'Wipe-out', and demodulation in a receiver is analogous to it. The theoretical treatment of demodulation is not simple, but the effect can be simply stated. The modulation of a weak but quite audible signal will become inaudible if a much stronger carrier be applied to the same detector. Since reaction provides a means whereby most signals can be made of far greater carrier amplitude than interfering ones, it reduces the actual interference heard considerably more than would be expected from inspection of the actual resonance curve. The effect cannot always be produced if the interfering station is very strong, since now reaction may be insufficient to bring up weaker carriers to exceed that strength before self-oscillation sets in, and in any case a sufficient ratio between wanted and unwanted carrier amplitude will not be obtainable.

Yet another phenomenon termed the 'Zeehen Effect', prevents the effective use of reaction in such a case. It is found that the presence of a strong oscillation upon the grid of a regenerating or oscillating valve can produce 'locking', the frequency at which regeneration occurs being pulled into synchronism with the strong oscillation. Thus it is found impossible to amplify a weak signal by reaction in a simple detector circuit if a much stronger local signal is present on an adjacent wavelength. The circuit is pulled into synchronization with the local signal, and there is no opportunity for demodulation to take place. The Zeehen effect is often useful in the laboratory. By its use an oscillator of poor frequency stability can be locked to the frequency of another more stable oscillator.

The superheterodyne overcomes those defects inherent in straight receivers at the expense of somewhat greater cost and complexity. Early types found it necessary to employ large numbers of valves, from eight upwards, to obtain even moderate

results. Improvements in valve efficiency, together with the production of the special dual types mentioned in Chapter V, have largely removed this difficulty, and to-day the number of stages may only exceed the straight receiver by two or three, a three-valve superheterodyne being possible. As a result, nearly all commercial receivers and a large proportion of others employ this principle.

We have examined the superheterodyne in outline in Fig. 112, and it remains now to see what benefit we can derive from the apparently unnecessary change to an intermediate frequency. This arises entirely from the convenience of carrying out the bulk of amplification at a fixed frequency. So great are the advantages of this that the losses unavoidable in heterodyning are worth while. From the practical point of view, it is easier to obtain high amplification per stage with stability, and to obtain the theoretical selectivity from band-pass filters or similar couplings, if these can be permanently tuned to one frequency. No change in performance will occur whatever the signal frequency, provided that this is not so high as to make the heterodyning process difficult, and the serious difficulty of ganging a number of circuits accurately over a wide frequency range does not occur. There is thus no limit to the selectivity which can be built into the I.F. amplifier by the use of an ample number of circuits, exactly tuned.

The greatest gain, however, occurs because a low I.F. can be chosen, generally lower than all or most of the incoming signals. It has been shown that fundamentally the ability to separate adjacent channels increases as frequency is lowered, and the sideband becomes a larger proportion of the carrier frequency. Amplification is also more easily obtained at low frequencies, where the effect of stray capacities is small. It is in fact possible to carry this change too far, for if an I.F. much below 100 Kc. be used, difficulty may be found in keeping sufficient band width to accommodate a wide sideband. A favoured I.F. until recently has been from 100 to 175 Kc., at which four circuits only will yield excellent selectivity. American broadcast receivers, which are not called upon to receive signals below 500 Kc., made use of 175 Kc. for many years; whilst in Europe where broadcast signals extend down to 150 Kc. (2,000 metres, 175 Kc. being about the frequency of Radio Paris), a lower I.F. near 100 Kc. was used. Commercial receivers naturally select

an I.F. which will meet their particular requirements, bearing in mind that lower frequencies give the best selectivity and are ideal for telegraphy, and that the I.F. should be outside the range of frequencies over which the receiver is to tune. It should also not coincide with that of a local transmitter or group of strong signals, since direct pickup of interference by the I.F. amplifier is not impossible. Screening is used to prevent this, and sometimes a wave trap tuned to the I.F. is inserted in the aerial lead to keep signals out.

The one drawback to very low I.F. is in the increased pre-selection necessary to keep out second channel interference, and were it not for this very low figures would be almost universal. To cheapen receivers it is a great help if pre-selection can be reduced, and mainly for this reason higher intermediate frequencies have come into use. From 400 to 500 Kc. is popular, since this lies between the long- and medium-wave broadcast bands, and is at the same time convenient for short-wave work. At this frequency it becomes necessary to use rather more tuned circuits to give high selectivity, six being a not unusual number. Inductances fitted with Iron dust cores, such as Ferrocart, have been developed to help the selectivity of individual couplings. The increased permeability of these cores enables a given inductance to be reached with fewer turns, which can be of thicker or stranded wire of reduced resistance. This lowered resistance raises the dynamic resistance of the coil to a greater extent than the increased core losses reduce it, resulting in a lower loss circuit of sharper resonance. Four such circuits arranged as two band-pass filters can practically equal the selectivity of six air-cored circuits, although the curve may be somewhat less rectangular in shape.

An I.F. of 450 Kc., or the still higher frequency sometimes used of about 1,600 Kc., lies higher than the incoming signal range. At one time this was considered as a different system, and was given the name of 'infradyne', superheterodyne being confined to the case in which the I.F. is lower than the signal frequency. For a signal at A Kc. and an I.F. of X Kc., the heterodyne oscillator can be set at $A+X$ or $A-X$ Kc. This is clearly only possible when A is greater than X, since otherwise $A-X$ becomes negative, and we can attach no meaning to a negative frequency. Hence when the I.F. is higher than the signal, only the summation frequency $A+X$ can be used for

the oscillator, the second channel being non-existent. This is an invaluable feature of infradyne operation, since although some pre-selection will still be desirable to guard against harmonic effects, it can be reduced to very small proportions. It can be readily understood on these grounds why an I.F. as high as 1,600 Kc. may be used, in spite of the difficulties in obtaining adequate selectivity at this frequency, which usually needs three stages of I.F. amplification. Of recent years there is a tendency to design broadcast receivers for 'all-wave' operation, a delightfully incorrect expression which often means that one or two short-wave ranges between 10 and 100 metres are added. All wavelengths cannot of course be received on any existing receiver, the nearest approach being a commercial model having interchangeable coils, which might cover the range from 1 to 30,000 metres. 'All-wave' sets and those designed specially for the very short wave-bands will tend to use a higher I.F., partly because the reduced effectiveness of pre-selecting circuits at high frequencies makes it difficult to discriminate sufficiently against second channel unless this be many Kc. from resonance, and also because the difficulty in adjusting such receivers may make it easier to employ somewhat less pre-selectivity. There are certain cases in short-wave operation when an altogether wider sideband will be needed, as in television, and this is best done if the I.F. be considerably raised.

Interference and Sideband Splash. With a superheterodyne receiver using a low I.F. there is no difficulty in obtaining a response curve about the width of one channel, or even less, although the slope of the sides of this curve will not be quite as steep as the ideal. Selectivity is now limited by the physical considerations of sideband response already outlined, and it becomes possible to check these in practice, and to consider whether any further improvement is possible. To understand the work which has been done in this direction, we must analyse the interference which can occur between two adjacent telephony transmissions. This can be regarded as the sum of the following effects:

1. The heterodyne beat between the respective carrier frequencies, usually an audible note of 9 or 10 Kc.
2. The possible direct reception of the sidebands of the unwanted station when tuned to that wanted.

3. Noises covered by the general term 'sideband splash' caused by the heterodyning of sidebands from the unwanted station by the carrier of the wanted station, and vice versa.
4. Heterodyne effects of the same kind due to heterodyning of the respective sets of sidebands against each other.

It is very instructive to consider these four components of interference individually. Firstly, there is the heterodyne beat

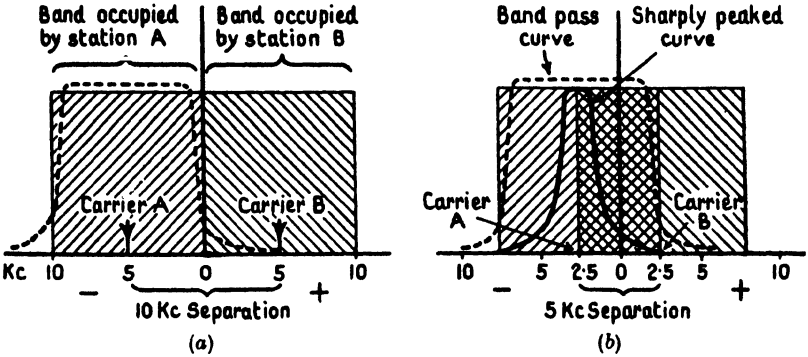


FIG. 121.

between the two carriers. This will seldom be zero, since there will be slight 'skirts' to the response curves of any practical receiver which will bring in a small potential from the unwanted carrier; but whether selectivity be of the band-pass or single peaked variety it should be possible to reduce this beat to a low intensity. The amplitude of any heterodyne beat is given by the product of that of the two oscillations which produce it. Thus since the wanted station carrier is tuned in at full strength, a very small residue from the other station can make itself heard. It is, however, possible to eliminate this beat in several ways, such as a low-frequency filter peaked to reject that particular frequency, popularly termed a 'whistle filter'; or by a rejector circuit at radio-frequency. It is thus safe to say that it need not exist in a first-class modern receiver.

Similar remarks refer to the second type of interference also. On the assumption that each station modulates up to a maximum sideband of 5 Kc. only, and that the carriers are 10 Kc. apart, then a good band-pass characteristic 10 Kc. in total width should completely cut out the unwanted sidebands. This is illustrated in Fig. 121 (*a*) where the two shaded areas represent

the frequency range covered by each station, over one of which is sketched a somewhat idealized band-pass response curve. Tests with a first-class modern receiver will show that interference of this type is in fact absent, and even in straight receivers the demodulation effects brought about through reaction can eliminate it if neither station be too strongly received. This interference can therefore also be considered as overcome in a good quality set.

Taking the fourth class of interference next, it will be obvious that the higher sidebands of each station will be able to heterodyne with each other, being little separated in frequency; but those from the unwanted station will be much weakened by selectivity, whilst those of the other we have stated do not in general carry much energy. The interference is thus a heterodyne between two small quantities, each of them transient, and will thus be a second order transient effect, which can be very largely neglected. Class three will similarly be impossible if the band-pass curve is perfect, and will only occur in proportion as the 'skirts' of this curve may allow slight response to the unwanted sidebands. Only the highest of these should be heard appreciably, and since their energy is again small in the case of speech or music, the interference will be slight. The low-frequency sidebands of each signal are separated by nearly 10 Kc., and should be unable to heterodyne appreciably at that separation unless exceptionally strong. For a *perfect* band-pass or a *perfect* peaked response curve no interference of types 3 or 4 can occur, and it will only make itself apparent as the response curve deviates from a complete cut-off at 5 Kc. each side of resonance. Thus it is clear that under the conditions stated either form of selectivity is theoretically capable of separating the two signals *completely*, the peaked curve differing only in poorer reproduction of the higher modulated frequencies. Since in practice, however, transmissions are modulated by frequencies higher than 5 Kc., 'splash' caused by the higher sidebands is often very objectionably evident.

It would be most desirable if transmissions could be separated by less than this limit without too great interference. Let us consider what would happen if the same two transmissions were moved together in frequency till their carriers differed by only 5 Kc., as shown in Fig. 121 (*b*). This also corresponds to the case of two carriers modulated up to 10 Kc., and differing by 9 or

10 Kc. in frequency, as occurs in actual broadcast reception. Clearly there is a complete overlapping of the sidebands of both stations within the doubly shaded area, and if the band-pass response curve be used separation is quite impossible. The use of a sharply peaked curve as shown by the full line also permits interference, but in practice a better separation is usually obtained for the following reasons. It will be seen that in the case of a band-pass filter the unwanted carrier *B* falls wholly or largely inside the curve. It is therefore strongly received. Since the energy of the rectified signals is proportional to the product of carrier and sideband energy, both the unwanted programme and heterodyne interference effects will be a maximum. In the case of a peaked curve, however, the carrier *B* falls at a low point near the 'skirts' of the curve, and is weakly received. Consequently, its heterodyne products including the unwanted programme will be weaker than in the former case. Better separation will have been achieved, but at the expense of poorer reproduction.

It should be pointed out in passing that if single sideband transmission or reception be employed it once more becomes possible to avoid interference. If, for example, only the left-hand sidebands shown were radiated from both stations, there would be no interference within the shaded area. Similarly, if only the left-hand sidebands are received, by a suitably shaped response curve, it will be possible to avoid interference from a transmission on one side of that being received, but not from both sides at the same time. Single sideband transmission has been referred to previously and in commercial applications it is frequently used, often with reduced or suppressed carrier. For broadcasting, however, the difficulties in design of satisfactory single sideband receivers and to a lesser extent of transmitters has so far hindered this major technical advance.

Stenode Reception. Up to about 1930 it was thought best to employ band-pass reception in high class equipment. It was realized that the lost modulation response resulting from a sharply peaked curve could be compensated by increased low-frequency amplification of the higher modulation frequencies, but no advantage could be seen in using this method. It was argued that such compensation would bring back fully all interference which the selective circuit had reduced, giving the equivalent effect to a band-pass response.

Dr. James Robinson was the first to point out that this is not the case, and that a very real reduction in interference takes place when a sharply peaked resonance curve is followed by tone correction. It was not difficult to demonstrate this improvement in the laboratory, but for a time its explanation remained obscure, and the proposal was received with ridicule by many authorities. Robinson was able to show that the demodulation of a weak carrier by a stronger one provides an adequate explanation, and was the first to point out the value of this effect. The use of a sharply peaked response has the effect of increasing the carrier of the wanted station, whilst that of an adjacent transmission remains very small. For example, a resonant circuit of very low damping will build up the voltage of a carrier in resonance to which it is tuned by a factor approximately Q times the voltage induced into the circuit. The carrier may be increased till it is many hundreds of times stronger than that of adjacent transmissions, and thus conditions are very favourable to demodulation. Once the modulation of interfering signals has been reduced through this effect, the amount of tone correction necessary to restore quality will not be sufficient to restore the interference to its original level.

We have seen that reaction is not entirely satisfactory in assisting demodulation, on account of the Zeehen effect. Also it is not easy to introduce sufficiently selective circuits at incoming signal frequency if these are to be adjustable, although they can be employed successfully when tuned permanently to a particular transmission. It is therefore convenient to provide the sharply peaked response in the Intermediate Frequency amplifier of a superheterodyne, where it can be readily obtained and tuning adjustments are fixed. It can be shown that if a single selective circuit of low damping be used, complete tone correction is provided by a low-frequency amplifier so designed that its gain increases proportionally with frequency up to the highest limit it is desired to reproduce.

The validity of this principle is now fully established, largely as a result of investigations carried out by the National Physics Laboratory.[1] It is sometimes referred to as the 'Stenode' principle, after the trade name of the receivers first brought out in

[1] Special Report No. 12 by F. M. Colebrook, published 1932 by the Radio Research Board.

this country, or as the principle of high selectivity followed by tone correction. It is employed to a limited extent in most modern commercial receivers, since it has become usual to introduce the maximum selectivity that practical considerations allow, without necessarily any attempt at a band-pass type of curve, and to make up for any resulting loss in the higher modulation frequencies by a rising curve in the low-frequency stages. We shall see that an uncorrected pentode forms a very simple way of getting such a characteristic in the output stage, whilst matching the loud-speaker impedance to that of the output stage at a high rather than a low frequency is another method which assists. Provided that the correction is not carried too far, a decided cheapening in the receiver together with better selectivity results. When pressed to its limit, however, the principle is less easy to apply. Additional valves are found necessary to provide the extra amplification at high modulation frequencies required for tone correction.

Crystal Filters. Robinson also introduced the quartz resonator as an ideal selective device. It has been shown in Chapter VI that this should be most suitable, since the equivalent damping is much less than any single circuits composed of inductance and capacity. It is quite possible to reach an attenuation of 10,000 times at 10 Kc. from resonance when quartz is used as a coupling element in an I.F. amplifier. This idea was very successful in achieving selectivity, and was at once seen to be very valuable in telegraphic reception, where only a narrow sideband near to the carrier frequency is required, and tone correction need not be very great. American engineers were the first to appreciate the commercial value of this new device, and under the title of the 'quartz crystal filter' it has become standard in most American Communication type receivers.

The basic circuit of a crystal filter used as an I.F. coupling is shown in Fig. 122. The I.F. stage V_1 is followed by an anode winding which forms the primary of a coupling transformer, and which need not be tuned. This is closely coupled to the secondary winding AB, centre tapped to 'earth' at C. A quartz resonator joins one side of this secondary winding to the point D, whilst the other end of this winding is taken to the same point through a small variable condenser K, termed the balancing condenser. The output potentials from this filter occur between the point D and earth, D being taken to the

grid of the following valve V_2, the grid circuit being closed through the impedance Z. When we are concerned with telegraphic signals, or desire maximum response to the carrier frequency, it may be an advantage to replace Z by a step-up transformer, thereby matching the low impedance of the crystal at resonance to the higher impedance of the grid V_2. The two portions AC and BC of the coil AB, together with the condenser K and the capacity of the crystal holder form the four arms of a Wheatstone bridge network, and if $AC = CB$ in

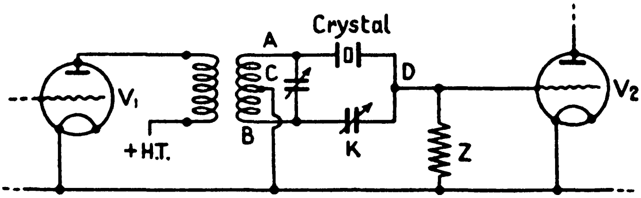


FIG. 122. Simple Crystal Filter.

inductance, then the bridge will be balanced (neglecting the effects of circuit resistance or losses) for alternating potentials when the capacity of K equals that of Q. No energy can then flow through the holder capacity, and it is possible to ensure that energy reaches the grid D through the quartz resonator only.

This effect can be visualized in another way if we realize that A and B are in opposite phase at all times, provided the coil AB is tuned to resonance. Thus the potential reaching D through the crystal holder capacity will be in opposite phase to that through K, and if adjusted to be equal, these two potentials will cancel. What is more important, however, is that by slightly altering the adjustment of K, a residual potential can be allowed to reach D in either phase as desired, and this will combine with the potential through the actual quartz resonator to modify the resulting response curve. The potential across a quartz resonator will reverse in phase as resonance is passed, just as in any other resonant circuit. Hence in a peaked curve such as that of B in Fig. 118 the phase of most of the left-hand side will be nearly opposite to that at a corresponding point on the right-hand side. This means that a potential applied through K will raise the response on that side of the resonance curve with which it is in phase, but will depress it similarly on the opposite side. This is valuable in two ways. Firstly, it

enables the quartz resonance curve to be made symmetrical when desired, the natural curve being unsymmetrical. Secondly, by adjustment of K a potential can be applied to D which is equal and opposite in phase to that existing at any selected point on the resonance curve. In this way the response at that point is depressed very nearly to zero, an attenuation of 60 to 80 db being possible. This means that by the adjustment of

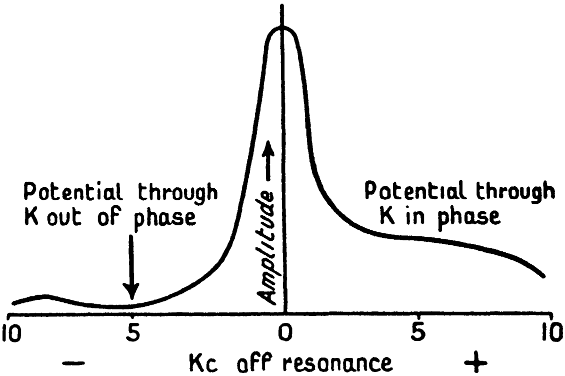


FIG. 123. Effect of Balancing Condenser.

K any single interfering carrier can be virtually eliminated, or any single heterodyne beat removed from the wanted signals, a most useful aid to selectivity. Alternatively, by adjusting K to a mean value which will reduce the response on one side of the resonance curve considerably, simultaneously raising the other, a very good approximation to a single sideband response curve results. This condition is illustrated in Fig. 123, the actual response curve being much sharper than that sketched. A station to the left of this curve will be very largely eliminated, whilst one having a carrier in resonance and sidebands on the right only will be received.

The tuning of the winding AB is found to exercise a pronounced effect upon the actual width of the response curve. This is widest and least selective when the circuit is exactly tuned to the crystal frequency, but becomes as sharp as that of the crystal itself when the circuit is widely detuned. This has been incorrectly described as varying the 'Q' of the crystal, but is in fact only an apparent variation due to changed conditions of loading, the crystal being physically unchanged. A 'variable width crystal filter' so achieved may possess an

effective response of only 200 cycles at the sharpest position, widening to some 5,000 cycles when the input circuit *AB* is in tune with the crystal.

The crystal filter is a very effective device for the reception of telegraphic signals through serious interference, provided that the transmission itself is stable in frequency. It is widely used for that purpose. Telephony can also be received through it, using tone correction, but here the engineering problems are a little less simple. The selectivity of a crystal is so high that it becomes difficult to adjust the receiver by hand, whilst incorrect tuning leads to imperfect tone correction and distortion. Automatic tuning which we have mentioned in an earlier chapter was introduced by Robinson to overcome this difficulty. There are also less vital difficulties which have prevented the wide adoption of the crystal for broadcast and general reception, but these are steadily yielding to research. Much has been done in recent years to improve both the crystals themselves and their methods of mounting, so that the device is fast becoming a reliable component, and it is likely that such improvements will lead to a far wider use of the crystal in the immediate future.

No discussion on selectivity would be complete without reference to one other serious problem, that of cross-modulation. This has already been alluded to, and as it occurs in the earliest stages of a receiver, no selectivity taking place later (such as in the intermediate frequency amplifier of a superheterodyne) can overcome it. Suppose we wish to receive a weak signal, and a very much stronger transmission exists on an adjacent frequency. The actual separation between these two carriers is unimportant, the only consideration being that the selectivity of those circuits before the first valve shall not be sufficient to separate them. In practice it seldom can be, since these circuits are damped by the aerial. Then on the first grid there will be two oscillatory potentials, that of the wanted station and that of the unwanted one, the latter being much the stronger. If the characteristic curve of this valve were strictly linear, the two sets of signals would be amplified without interaction. Unfortunately all practical valves show slight curvature, with the result that slight rectification is inevitable. Directly this occurs the two signals are able to modulate each other. The stronger signal sweeps the grid potential in accordance with its own

modulation, and thus varies the point upon the curve at which the weaker wanted signals are being amplified. Since the curve is not a straight line, but varies slightly in slope, the amplification of the weaker signal becomes controlled by the stronger, and is thus modulated by it. Cross-modulation can also be regarded as modulation of the weaker carrier by low-frequency grid potentials, produced by the partial rectification of the stronger signal. Traces of the unwanted modulation then occur upon the wanted carrier, and since these are in the form of actual modulation, they cannot be separated by any subsequent selective process.

Cross-modulation is a valve defect, and as research on valves is leading to increasingly linear characteristics, it may be expected to disappear in due course. In the meantime, some measure of this defect occurs in the early stages of all receivers, but it only becomes important when selectivity is of the highest order. In this case it may be found to mar performance appreciably, when the next channel to a local broadcasting station is wanted for example. Complete separation under these conditions is often impossible. Cross-modulation is reduced if selectivity can be concentrated before the first valve, but it has been explained that this is far from easy, except in the case of receivers permanently tuned to a particular transmission. It is good practice, however, to employ as much selectivity immediately following the aerial as circumstances will allow. Valves of high-alternating current resistance are the most prone to cross-modulation, since their characteristics are less linear, and smaller amplitude can be accepted on the grid before rectification becomes serious. Similarly, the use of high stage gain in a receiver encourages this defect, and should be avoided in the early stages.

Our conclusions on selectivity may be summed up as follows. The band-pass filter has long been regarded as the ideal selective device. It can separate transmissions perfectly if sidebands do not overlap, and results in a receiver which is easy to handle and to engineer, but better results are possible if a sharply peaked resonance curve be used. This attenuates high-modulation frequencies, which must be restored by a corrected low-frequency response. This system gives increasing separation as the damping of the selective circuit is reduced; but becomes increasingly difficult to apply. As a result most commercial receivers employ

some measure of peaked response and tone correction. Cross-modulation is a valve defect which can reduce the advantages of high selectivity, and must therefore be guarded against. This is best done by the use of maximum selectivity before the first valve. Demodulation in the detector assists selectivity, and should be encouraged by increasing the wanted carrier as much as possible, in which reaction is helpful. All selective circuits can most easily and effectively be applied at a fixed frequency, and the intermediate frequency amplifier of the superheterodyne is well suited to the purpose.

EXAMINATION QUESTIONS

1. What advantage is derived from using a coupled circuit in a receiver? How is the response of the receiver to incoming signals modified as the coupling is gradually increased from zero to a maximum? *City and Guilds Institute. Preliminary Exam. 1934.*

2. How would you define the following terms with reference to a tuned circuit: (*a*) Sensitivity; (*b*) Selectivity; (*c*) Fidelity? Is there any connexion between them?
Institute of Wireless Technology. June 1934.

3. What do you understand by the 'sideband theory'?

4. Explain why there is a limit to the selectivity of a receiver which is intended for the reception of telephony.

5. What are the several components which combine to make up the interference heard from an adjacent broadcast transmission in a receiver tuned to a neighbouring wavelength? Two telephony transmissions differ in frequency by 8 kilocycles. They are received on a particular receiver at equal field strength. What will be the highest modulation frequencies that can be reproduced from the loud speaker without appreciable interference from the unwanted transmission, if the receiver contains the following selective circuits:
(*a*) A perfect band-pass filter.
(*b*) A pair of flatly tuned circuits.
(*c*) A single circuit of very high 'Q', followed by low-frequency tone correction.
In what way will the results differ if one station exceeds the other in strength by 1,000 to 1? Consider the receiver to be tuned to each transmission in turn, and give the reasons for your statements.

6. What is meant by the 'Q' of a circuit? If a parallel resonant circuit consists of a condenser C in parallel with an inductance L

having a resistance R, show from first principles that the parallel reactance at resonance is Q times the reactance of one branch of the circuit. (See also earlier chapters.) *A. I. W. T. June 1937.*

7. Show that a modulated high-frequency current contains two additional frequencies for each modulation frequency. Explain how speech modulation of high-frequency currents is effected in wireless telephony. *Grad. I. E. E. 1935.*

8. Explain how superheterodyne reception assists the design of a highly selective receiver. What are the advantages of a high or low intermediate frequency from the point of view of selectivity ?

9. What is a 'quartz crystal filter' or 'crystal gate' ? Outline the action of this aid to extreme selectivity.

10. What is the meaning of the term 'detector demodulation', and is it of any use in the separation of wireless signals ?

11. What do you understand by (*a*) the Stenode principle, (*b*) cross-modulation, (*c*) sideband splash ?

CHAPTER XII

FAITHFUL REPRODUCTION IN RADIO-TELEPHONY

A CONTINUOUS wave modulated by music or speech can be received by any type of receiver. For radio-telephony to be successful, however, it is essential that the sounds produced at the receiving end should faithfully reproduce those at the transmitting end. To do this it is necessary that no distortion should be produced by either the receiving or the transmitting apparatus.

We have seen already that the quality of a musical note or sound is produced by the combination of a number of pure sine waves, the sum total of these producing the sound as we perceive it. After transmission, the relative phases of the various component waves making up the sound may be considerably changed, but the ear fortunately does not seem to be greatly affected by such changes in phase. What is essential, however, is that the relative distribution of energy between the various component waves should be unaltered by the transmission. This implies that the apparatus as a whole should be capable of functioning independently of frequency over the whole band of frequencies covered by speech or music. The low notes, for example, must not be lost or accentuated, nor must the intensity of the high notes be levelled down. Fortunately we can rely on a large amount of accommodation by the ear, which will give us an impression that we hear a good reproduction of the original sound, although actual measurement might show that it is as much as 20 per cent. out. Let us now review the various links in the chain by which sounds are transmitted and received by radio-telephony.

In Chapter IX we saw that the simple carbon microphone, whilst sensitive and satisfactory for intelligible speech, lacks the wide and uniform frequency range necessary for first-class musical reproduction. Much improvement is possible by the transverse current and similar constructions, but it is generally better to make use of one of the other types of microphone there mentioned. We have not the space to consider each of these in detail, and fortunately this is hardly necessary from the point of view of frequency response, since when correctly designed all

the types described are equally capable of handling the entire useful audible range. Many microphones will respond down to substantially zero frequency, and a lower limit of some 20 to 50 cycles can be safely assumed. Examples differ rather more at the high-frequency end of the scale, but it is a poor microphone which will not respond to well above 5,000 cycles. The majority respond up to between 10,000 and 20,000 cycles, a higher frequency than the remainder of the channel can generally handle. There is little to choose between the several principles of construction, but differences naturally exist between microphones of different makes, or by various designers. These mainly consist of minor peaks at particular frequencies, or a slight rising or falling characteristic at the extreme high and low frequencies. Broadly speaking, however, the microphone is no longer one of the weakest links in high-quality reproduction as were the earlier carbon types.

The carbon microphones responded to variations in air pressure caused by the sound waves, but in the more highly developed types it has become possible to distinguish two distinct classes, those which respond to air pressure, and others which respond to the velocity of the oscillating air particles. Any microphone having a diaphragm with an enclosed space behind it must belong to the former class, since changes of air pressure will react upon the front of the diaphragm, but cannot reach the back. A closed cavity behind a diaphragm is not ideal since it tends to introduce resonance at a particular frequency, and give rise to unequal damping. By scientific proportioning of the cavity, however, it can be made to form an acoustic impedance which plays a part in the correct working of the microphone. A narrow channel may be provided between this space and the outside air, which will allow a gradual equalization of pressure with changes of temperature, without allowing free access of sound-pressure waves to the rear of the diaphragm.

Any microphone designed on these lines will be a pressure-operated type. If, however, free access be available to both sides of the moving element, such as in the case of a ribbon stretched between the pole-pieces of a magnetic system allowing unimpeded air-circulation on all sides, then the motion of this light ribbon will tend to follow the changing velocity of the air stream in which it lies. It will not be affected by pressure changes, since these act equally on all sides of the ribbon. The

velocity microphone is a newer idea than the pressure type, and offers advantages in certain cases. This is not so much a matter of frequency response, which may be excellent for either type, as one of directional effect.

The carbon microphone, or any pressure microphone, responds best to sounds which come from the front of the diaphragm. It will of course respond to those from other directions, but unfortunately the frequency response differs considerably with the angle at which a sound arrives. This is not very important when a speaker is close to the front of the microphone, but becomes undesirable if the source of sound extends for a wide arc, as might occur in the case of an orchestra. It is most serious when sounds are to be picked up in a large studio having echo and reverberation, because echoes will reach a microphone from all directions. If the frequency characteristic differs markedly between echo and original sound, a very unnatural result is to be expected.

The velocity microphone can be far more satisfactory in this respect. The simple ribbon type has a definite directional characteristic, responding equally to sounds from before or behind the ribbon, but being quite dead to those arriving from the sides, because the ribbon is not free to vibrate in a lateral plane. Within the active area the frequency response is much more uniform, and thus the microphone responds more naturally to distributed sounds or to echoes. The presence of certain dead directions may be of great assistance in certain classes of work, such as sound-film recording, where there is a source of sound such as a film-camera which must not be picked up. The camera is placed along one of the directions in which the microphone has no response, and may then be quite near without any elaborate quieting arrangements being needed. In the case of earlier microphones the camera was used within an elaborate and inconvenient sound-proof enclosure, termed a 'blimp'. Many other cases will occur to the reader in which unwanted sounds can be reduced by a directional microphone.

For general work, however, the most natural result is to be expected if the microphones are entirely non-directional. This also makes it simpler to combine the output from a number of microphones. A combination of ribbon-velocity and pressure types has been evolved which is dead in one direction only. The 'polar diagram' of a microphone can be plotted, showing

in a horizontal plane the response to sounds at all angles of arrival relative to that from the front. These are similar to the radiation diagrams of aerials shown in Chapter XIV. Fig. 124 shows typical diagrams for four types, neglecting variations in frequency response. The cardioidal or heart-shaped curve of

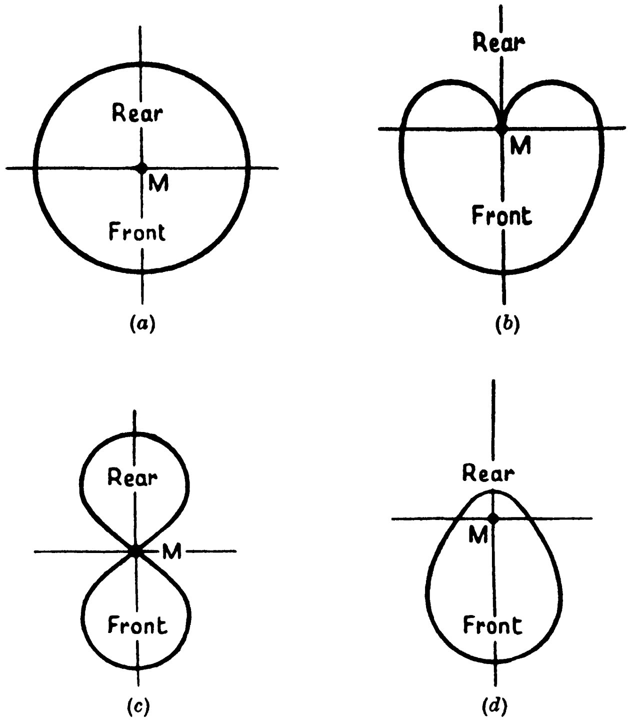


FIG. 124.

Fig. 124 *b* is the useful one just mentioned, whilst that of a truly non-directional microphone would be a perfect circle, as in Fig. 124 *a*. Fig. 124 *c* is the polar diagram of a ribbon microphone, and Fig. 124 *d* is typical of a carbon or condenser microphone of the earlier type.

It is possible to approach quite nearly to the non-directional ideal. The piezo-electric crystal microphone, employing one or more small crystals or 'sound-cells' open on all sides, can give a near approach to equal response from all angles. It may be placed in the centre of an orchestra or studio of performers and will give an excellently balanced reproduction. The General

Electric Company has recently developed a moving-coil microphone of similar properties. The result is made possible by careful design of the outer case, which is spherical and fitted with a baffle-plate to equalize the response from the front, which would otherwise predominate. It is desirable that microphones should be small and streamlined, so that they will distort the sound field as little as possible, and development on these lines has greatly improved their performance.

We see therefore that a modern microphone chosen for its particular class of work will pick up sounds with very little loss, a statement which was far from true of the earlier varieties. Unfortunately a penalty must be paid for this performance, since the oscillatory potentials produced by such microphones are very small in comparison with the carbon types. As an example, a good carbon microphone used with the correct exciting potential and impedance-matching transformer may produce an output of 60 millivolts per bar (one dyne per sq. cm.) pressure upon the diaphragm. This is quite sufficient to operate the first valve of a normal modulation amplifier, or for transmission over long cables without interference by external noises or stray potentials. The output from high-quality microphones is unfortunately very much lower, on account of the light moving parts and high damping that are used. Typical figures are from 3 to considerably less than 1 millivolt per bar. It is usual to express these outputs in decibels below a standard level of 1 volt per bar, in which case the carbon microphone may be less than 20 db. below, whilst the high-quality types vary from 45 to 90 db. below this level.

At outputs below 50 db. it is not advisable to pass the potentials through cables of considerable length, or to apply them directly to a modulation amplifier, as the risk of stray noise pick-up is too great. It is therefore usual to employ a 'pre-amplifier' of from one to three stages, of specially quiet and non-microphonic construction, and generally placed very near to or in the same case as the microphone itself. In this way the leads to the first grid are very short and well protected, and the signals are amplified to a safe level before passing into any long exposed cable. The pre-amplifier may be supplied from self-contained batteries or receive power along separate conductors included in the microphone connecting cable. It forms one unit with the latter, and is treated as part of the microphone

for practical purposes. Condenser microphones in particular are invariably fitted with a pre-amplifier, since it is necessary to connect the condenser directly to the grid of a valve, as shown in Fig. 91. The use of a cable at this point would introduce a fixed capacity in parallel with the microphone, which would seriously reduce its output.

In the amplification of microphone output potentials we are concerned with the strength of these relative to any extraneous potentials which may be picked up by the cables between microphone and radio-transmitter. These will of course be screened by metal casing but will still be affected by any magnetic fields through which they may pass. Now, any given quantity of stray energy induced into the cables will set up an interfering potential proportional to the impedance across which it exists. Thus a high-impedance line will be more subject to such troubles than one of low impedance. Both the average microphone and a valve grid circuit possess a high impedance, in excess of 10,000 ohms, and the circuit coupling them would thus be very subject to interference. It has been found impossible to work under such conditions, and the usual procedure is to change to a low impedance whenever the signals have to pass over long cables.

To do this, transformers are employed at each end of the line. At the input a transformer is used having a primary impedance equal to that of the microphone or pre-amplifier output. This input transformer has a large step-down ratio, the secondary being designed to match an impedance usually between 200 and 600 ohms. This now becomes the line impedance, to which the signal potentials have been suitably 'scaled down' by the transformer without loss of energy. We know that the condition for maximum energy-transfer is that the impedance of the source shall be equal to that of the load. This condition is maintained at each side of the input transformer, termed an 'impedance-matching' or 'line-input transformer'.

The signals are now at lower potential and correspondingly increased current level, and passing over a line of only a few hundred ohms impedance. Induced currents will be small relative to the signal currents, and will be of little account. It is usual to introduce all volume-control or 'fader' panels and mixing and control circuits into this low-impedance line, which can safely be taken anywhere and treated with a minimum of care. At the termination of this line a second transformer is used,

having a primary impedance equal to that of the line, and a high step-up ratio to match the grid impedance of the first modulation amplifier stage. The signals thus reach this amplifier in their original form, as relatively high potentials. When a very long line exists between studios and transmitter, it may be necessary to introduce additional amplifiers to restore the signal amplitude after attenuation by line resistance and capacity. Here again the change to a low impedance is necessary, for not only does it reduce these losses, but it prevents higher modulation frequencies from being by-passed through the capacity between the conductors of a long cable. The comparatively low reactance of this capacity at high frequencies would form a serious leakage path across a high-impedance line, but if the line be of low impedance the capacitive reactance may be neglected. There is, however, always some loss of the higher frequencies from this cause, and to compensate for this a rising characteristic may be imparted to the amplifiers by designing the intervalve couplings to operate with increasing efficiency at high frequencies. It is not unusual in a pre-amplifier to include correcting filters which will remove any resonant peak or known defect in the microphone output.

The microphone potentials have now reached the modulation amplifier, sufficiently free from background noise. Clearly all amplifiers must be designed to give minimum distortion by the methods outlined in Chapter VIII, and the modulation system must be linear, as shown in Chapter IX. Both these sections of the equipment can be very nearly perfect if correctly engineered, and will not in general contribute any serious distortion. We can assume therefore that a copy of the original sound-waves will be radiated from the aerial as modulation, differing in relative amplitude by perhaps 5 or 10 per cent. at most within the frequency range transmitted.

Distortion in Receivers. Transmission through the ether is inherently distortionless, and this will be the case within the ground-wave service area of a broadcasting station. At great distances selective fading may change the relative amplitude of sidebands and in some cases seriously impair the quality of reproduction. We have studied in the last chapter the effects which excessive selectivity may have upon quality within the receiver, leading to the attenuation of the upper sidebands and reduction in the higher frequencies contained in the

modulation. This must of course be avoided by the methods explained, and excessive interference from other transmissions prevented if possible. No further comment is now necessary on these points. Little distortion is to be expected from other causes within the high-frequency stages of a receiver, although a non-linear characteristic here, as at all other points, can modify the signal wave form and thus distort the modulation. The effect of this is to introduce high harmonics that were not present at the microphone, but the effect is usually very slight in the pre-detector stages of a receiver.

There are other ways in which radio-frequency stages can cause distortion, however. We have mentioned cross-modulation, which accompanies partial rectification in these stages. This condition can result in the negative fringes of the modulation envelope being amplified less than the positive fringes, with a resulting asymmetry of wave form. This can occur if the grid bias is excessive or the signal amplitude too great, and is an additional reason against volume control by reduced valve-operating potentials. Should bias be too low, so that grid current flows during the positive peaks, then the valve will become partially conductive during these moments and will damp the tuned circuit across which it is placed. The amplitude of deeply modulated passages may be reduced, giving rise to 'amplitude distortion' which also occurs when any portion of the receiver overloads or reaches a saturated condition so that further increase is impossible. Pre-detector stages should thus be correctly biased and as linear as the valves permit.

Of the remaining points where distortion is probable in a receiver the detector is amongst the most vital. We have seen in Chapter VII that detection is fundamentally a non-linear process, whilst linearity has been stressed as the key factor in distortionless reproduction. The earlier grid-and-anode type detectors were developed mainly with sensitivity in view and were admittedly imperfect. A perfect detector would possess infinite resistance to potentials of one polarity and zero resistance to those of opposite polarity. The diode is nearly an infinite resistance when its anode is negative, although it still possesses a capacitive reactance which cannot be neglected at the higher carrier frequencies. On the positive side the curve is of the type shown in Fig. 33, having a brief curved portion for very low input, a substantially linear middle portion, and a falling-off

portion as anode-current saturation is reached. The curved 'weak signal' portion should clearly be avoided, since both sensitivity and distortion will be at their worst. To minimize this we have already seen that detection should take place at high level, namely after considerable amplification. The upper limit of the curve must also be avoided, but this gives no trouble in practice if valves of ample emission are used.

Practical diodes also depart from the ideal in that anode current does not commence to flow at exactly zero anode potential, but at a slightly higher value. As a result, low values of modulation upon a carrier which reaches the linear central portion of a diode curve will be rectified with very little distortion. Should modulation exceed some 90 per cent., however, the carrier will approach zero during the troughs of the modulated wave form, and at these times there will be a departure from linearity. Distortion therefore increases when modulation becomes very deep.

When studying detection we saw that a resistance-capacity filter is necessary following a detector, to separate the modulation from the radio-frequency and direct-current components. This was shown as the condenser C_3 and load resistance R of Figs. 75 and 117. These necessary steps can also introduce a measure of frequency discrimination within the modulation range itself. The function of C_3 is to by-pass unwanted high-frequency components through the effect of its low reactance relative to R. This effect will persist to a reduced extent at the higher audible frequencies also, weakening them relative to the lower notes, at which the reactance of C_3 can be entirely neglected. To minimize this effect C_3 is kept as small as possible consistent with adequate filtering action, 0·0001 microfarad being usual at 450 kilocycles carrier; also R must not be too high. To obtain maximum output R would be a megohm or more, but this is only practicable when receiving telegraphic or low-quality modulation. To reproduce well above 5,000 cycles R should not exceed about a half-megohm. The slight attenuation due to C_3 at any frequency can be estimated on the lines given on p. 213 of Chapter VIII. When the highest quality of reproduction up to 10,000 cycles or more is wanted, R will be reduced to from 250,000 to 100,000 ohms.

Any detector stage places a load upon the tuned circuit across which it is connected. This is small in the case of an anode

detector, which is negatively biased and therefore passes no grid current. It is somewhat greater when grid detection is used, and generally still greater for the diode, for since the impedance of this is low during positive half-cycles the resistance R is then virtually across this circuit. In the latter two cases we cannot neglect the loading effect of the detector, which damps the tuned circuit considerably, reducing the amplification of the preceding stage and the selectivity of the circuit. Moreover, since the impedance of the valve is not quite independent of applied potential, this load will vary somewhat with depth of modulation, causing a degree of distortion. To minimize this the preceding stage should be regarded as a power-amplifier rather than a voltage-amplifier. It should pass sufficient anode current, and the anode impedance should be matched for efficient power output. There should be sufficient power to provide the current through the diode at positive peaks without appreciable drop in potential. Further to assist performance it is preferable to provide a step-down of impedance from that of the anode circuit to that of the diode. This may be done by the use of a transformer having fewer secondary than primary turns, or by tapping the diode across part only of the anode coil. An actual gain in potential across the detector may result from this. The step-down ratio may be more than offset by increased dynamic resistance of the anode circuit, resulting in better amplification by the preceding stage and greater potential across the primary winding. At the same time selectivity will benefit.

Following a diode detector operating at high level, we have seen that considerable low-frequency amplification is neither necessary nor desirable. There may be sufficient rectified modulation to drive a power-output stage directly, but more often a single amplifying stage will be provided between the two. This may be resistance-capacity coupled, and since the gain needed will not be very great a low-magnification valve will suffice. If we select a triode of fairly low impedance and amplification factor, the couplings can be chosen for minimum distortion rather than for efficient amplification, and the valve will easily accept the full available signal amplitude at its grid without departing from the most linear portion of its characteristic. Following the principles laid down in the chapter on amplification, the distortion introduced by this stage should be very small indeed. It is in the following power stage and loud speaker

that the greatest departure from the ideal is likely to occur, and since the design of these units is essentially determined by considerations of quality they have been left for treatment in this chapter.

When discussing amplification we assumed as our objective the reproduction of an oscillatory potential on an enlarged scale and with identical wave form. No considerations of power entered into this conception, and although a potential change can hardly occur in practice without an associated expenditure of energy, however minute, we have generally assumed this to be negligible both at the input and at the output of an amplifying valve. Actually this is only approximately true of the grid circuit, on the assumption that no grid current flows, and neglecting the grid-cathode impedance, which is not infinite. If this load is not negligible, power must be provided in order to maintain the desired oscillatory potential across it, and this power is drawn from the anode current of the preceding stage. Thus anode-circuit power considerations can seldom be neglected, and several times it has been necessary to refer to the matching of anode impedance so that efficient power-transfer will take place. This aspect of the matter has been kept in the background as far as possible, in order not to confuse the treatment of voltage amplification. If our object is to maintain constant voltage amplification at high frequencies, when there is a comparatively low shunt impedance due to stray capacities or other sources of loss, the problem is really a power one. Sufficient power must be available to supply that which will be dissipated in the shunt impedance. Unless this is attended to, there can be no possibility of maintaining the desired oscillatory potential across it. It is partly the increased power loss occurring in dielectrics and therefore in most component parts of an amplifier which makes amplification at very high frequencies difficult.

Power Amplification. An oscillatory anode current of a few milliamperes flowing through a high impedance represents quite appreciable power, and for this reason most amplifying stages are both potential- and power-amplifiers. There will be few cases in which the power developed across the anode-circuit load does not exceed that supplied to the grid circuit. The only basic difference between a voltage- and a power-amplifier is thus the fact that in the second case conditions of operation are chosen to favour the maximum output of power, whilst in the

first they are chosen for reasons of amplification only, power output being a secondary consideration. Moreover, since a valve will generally deliver far more power if distortion be permitted, as in the class C stages which we noticed were useful in transmission, the present problem centres on the selection of correct working conditions for the maximum undistorted power output. A power-amplifier can be defined as a valve so adjusted that the maximum power output is obtained having a current wave form identical with that of the potential wave form applied to its grid. It is thus a form of current-transformer.

In any amplifying valve variations of grid potential give rise to similar variations of anode current. In the case of a voltage-amplifier these are translated back into potential changes by passing them through a high impedance Z, which for best amplification will be as high as other considerations permit. In the power case, Z is selected so that maximum power will be developed therein, which will occur when the impedance equals the mean alternating-current resistance of the valve. This is the condition for maximum distortionless power output over the whole range of audible frequencies, which may be taken as from 20 to 10,000 cycles. It is independent of the characteristics of the valve used, except for the actual value which must be given to Z in relation to the R_a of that valve.

The simplest form Z can take will be a resistance which is essentially independent of frequency. Since, however, it will be shown that a power valve must pass a considerable steady anode current, this will give rise to a large potential drop across the resistance, necessitating an excessively high-voltage anode supply. This argument was given when treating resistance-coupled amplifiers, but in the present case it is even more vital. Economy will not allow any expenditure of energy which does not play a useful part in the working of the stage, and energy lost as heat in the anode resistance comes under this heading. For this reason the impedance Z nearly always takes the form of an inductance, such as an iron-cored choke coil of small ohmic resistance. It must not be treated as a purely inductive reactance, however, since the resistance and self-capacity are seldom negligible.

The early power-output stage was a simple triode, provided when necessary with increased filament emission and larger anode, so that larger anode currents could be safely employed.

Even to-day this remains the best arrangement from the point of view of distortion, so it will be as well to review its chief properties. The power triode will be biased to the class A condition specified in Chapter VIII, namely, to the middle of the straightest portion of its characteristic curve. A comparatively high negative bias is needed to do this. Therefore grid current will not flow, and extremely little driving power is demanded from the preceding stage. Should grid current flow during the maximum positive peaks, distortion is to be expected, both from unequal grid loading and because the linear part of the curve will have been exceeded. This is the overloaded condition, which must never be allowed to occur in a class A biased stage.

The exact determination of best operating bias is not so simple as the above approximate rule would suggest. If we examine a set of valve characteristic curves such as those of Fig. 38 it will be seen that the mid-point of the straightest portion lies near to zero bias; and were this value to be used in practice a quite moderate oscillatory grid potential would make the grid positive during peaks. The maximum oscillatory potential, or 'grid swing', to use a convenient popular term, would not provide anything like the full power output of which the valve is capable, for to get this the anode current must vary over the widest possible linear limits. The highest grid swing possible without encroaching upon the positive region will clearly be obtained if the grid be biased midway between zero and the negative potential at which anode current becomes zero; but if this is done it would seem that we shall be operating over the curved lower portion of the characteristic, and distortion will result.

The explanation of this apparent disagreement lies in the fact that the anode current-grid bias curve, which we have been treating up till now as the basic valve characteristic, is not adequate to meet the present case. It is termed a 'static' curve, because it is drawn from readings of bias and current taken whilst the valve is in the steady state, with no resistance in the anode circuit. Under oscillatory conditions and with an anode load, however, the conditions are quite different. As anode current increases there will be a fall of anode potential. Any change in momentary grid potential will give rise to a change in both anode current and potential simultaneously, whilst in plotting the static characteristic the anode potential has been

assumed constant. As a result of this variation, the anode potential when anode current is low may rise above the mean anode voltage. The maximum in the case of a triode is double the latter figure. This will have the effect of extending and straightening the negative end of the static curve, for in effect the valve operates with higher anode voltage as the grid be-

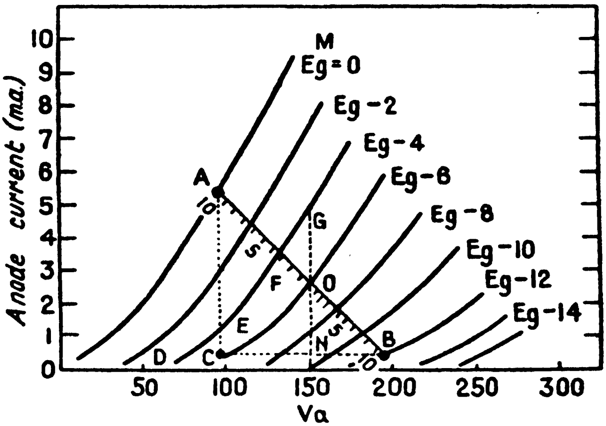


FIG. 125.

comes increasingly negative. Hence the symmetry is better than it appears from the static curves, and a new set of curves termed 'dynamic characteristics' can be prepared in which this effect is taken into account. From these the true grid bias corresponding to a symmetrical operating-point could be read off.

Static curves are, however, more easily measured, and if they are to be used a more useful one is that showing anode current plotted against anode voltage. A typical family of these is shown in Fig. 125, for various values of fixed grid bias. Let the point O be that at which we decide to bias the grid to fit in with the considerations which follow. Then a line AB can be drawn through O as shown. This line is an axis of varying grid potential, and so AB can represent the maximum grid swing which can be applied before the grid becomes positive, beyond A. Now the slope of AB measures a resistance, because it represents a constant rate of change of current with potential. The actual value of this resistance is BC/AC, and it represents the resistive load in the anode circuit to yield the conditions represented

by the family of curves. The power expended in this load under oscillatory conditions is proportional to the area of the triangle ACB, and so if the 'load line' AB is chosen to make this a maximum, the load which it then represents will be the optimum load for maximum power output.

Assume that the load line has been drawn through O to meet this requirement. For symmetrical operation with low distortion O must be chosen so that $AO = OB$ within at least 10 per cent. Moreover O must lie at values of steady anode voltage and current which are safe for the valve, and which when multiplied together, to give the number of watts dissipated as heat at the anode, will not exceed the makers' rating. These considerations fix the position of O definitely. The graphical construction of a load line is a most useful one when designing power-amplifiers. For example, when the resistance or impedance of the load is known, we can draw the corresponding load line, and from this can find at once the best working potentials for any given type of valve.

If the optimum load be estimated from the slope of AB, or calculated from the dynamic curves, it will be found to be about twice the alternating-current resistance R_a of the valve. This may seem a direct contradiction of the principle that maximum power is delivered into any load when the resistance of this equals that of the source. It is not actually so, for the principle is stated on the assumption that both resistances are constant, whereas in the present case they are not. The effective dynamic impedance of the valve will be a quantity that varies with oscillatory grid potential, and it is because of this that the optimum load does not equal the static value R_a.

Assuming now that the optimum load R is about twice R_a, it is possible to reach further conclusions regarding the efficiency and best form of anode impedance. Consider firstly the amplification possible. This is given by $\frac{R}{R+R_a}\mu = \frac{2}{3}\mu$. So the voltage amplification of any triode used as a power amplifier will be about two-thirds the amplification factor of the valve. For an oscillatory grid swing of X volts, peak to peak, the oscillatory anode potential across R will be $\frac{2}{3}\mu X$. The mean current through R is given by this potential divided by R; and the power output into R by current squared times R, or ${}^{4}\mu^2X^2/R$. Once again it is evident that for maximum power

amplification R should be low and μ high, and, remembering that $R = 2R_a$, a valve of high mutual conductance is needed.

Power Output. The maximum undistorted power output is a vital quantity to know, since it determines the volume of sound that can be obtained from a given loud speaker with minimum distortion. It depends, however, upon the length of the straight portion of the valve characteristic, and is therefore variable from one type to another. A definition of 'undistorted' is also necessary, since some distortion is invariably present. The most usual definition is the maximum power obtainable from a sinusoidal input potential at some specified frequency, such as 1,000 cycles per second, with a total content of not more than 5 per cent. of second harmonic introduced into the output wave form.

We cannot calculate the exact undistorted output without making use of the actual dynamic curves of the valve employed. For average triodes, however, experience has shown that the ratio of useful power output to total power input is about 1:5, approaching 1:4 for the latest valve types. The efficiency of the stage is thus poor, lying between 20 and 25 per cent. only, without considering further loss in the ohmic resistance of the load. To fix an idea of scale, the voice of a speaker upon a platform averages a power of about 0·25 watt, reaching perhaps 1 watt when the voice is raised. The efficiency of most loud speakers is low, a good figure for a moving-coil type used without a horn being 10 per cent. Hence an acoustical output of a quarter-watt representing speech at natural level would require an electrical power output from the amplifier of ten times this, namely 2·5 watts. For the reproduction of music it is considered necessary to provide more energy, and an output of 10 watts would not be considered excessive if loud passages are not to exceed 5 per cent. first-harmonic distortion. Taking an average output as 5 watts, typical of the better-class broadcast receivers, this implies a total anode-circuit input of from 20 to 25 watts to a triode power stage. At a high-tension voltage of 500, this implies an anode current of 40 or 50 milliamperes, or at 250 volts as much as 100 milliamperes: no inconsiderable current if provided by batteries. This also assumes no resistive loss in the anode impedance. A suitable value of resistance would be 4,000 ohms, typical of the alternating-current resistance of suitable valves. At 50 milliamperes this will drop

200 volts, or 10 watts of energy, roughly half the total input. The advantage of using an inductive impedance is therefore obvious if consumption is to be kept down.

We must now turn to the consideration of practical power-output circuits in which an inductive load is used. Fig. 126 shows the two most important cases. In the first case a choke

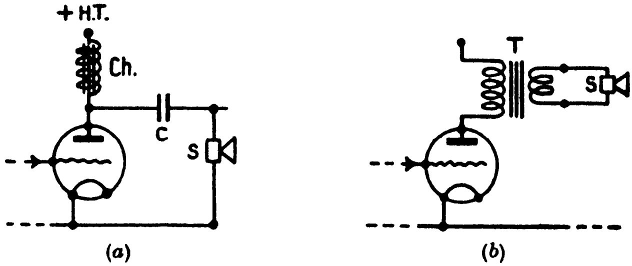


FIG. 126. Loud-speaker Coupling Circuits.

Ch is inserted directly into the anode circuit, and the potentials built up across it are transferred through the large condenser *C* to the loud speaker or other device which we want to drive, shown at *S*. Let its impedance be Z_s. Now the condenser *C* serves merely to keep the anode current from *S*, and will be of from 2 to 10 microfarads capacity. Its reactance will therefore be small at average frequencies, and for the moment will be neglected. With this assumption the choke *Ch* and true load *S* will be in parallel between the valve anode and cathode, and will share the output energy between them. Their joint impedance must thus equal the optimum load *R*. Except at very high frequencies we can regard them without serious error as mainly inductive reactances, when clearly the most energy will reach the speaker *S* if the reactance of *Ch* is high compared with Z_s. This is the case in practice, *Ch* being as large as possible.

The whole reason for using this apparently wasteful parallel arrangement, or 'choke-condenser filter', as it is termed, is to provide a low-resistance path for the anode current through *Ch* whilst keeping it out of the loud-speaker windings. The choke can be comparatively bulky, and it is not difficult to design a choke having high reactance, low ohmic resistance, and able to carry large currents without damage or magnetic saturation of the core. The speaker, on the other hand, is limited in size from other considerations, its windings may be of very fine wire and

may be damaged by large currents, while its inductance may be reduced by magnetic saturation.

Where quality is less important, or when a large and suitable speaker is used, it is quite possible to insert it directly into the anode circuit, as was done in earlier circuits. The impedance of many 'high-resistance' speakers is about equal to the optimum load of small power valves, and in such cases there is no serious objection to so doing. When anode voltages are high, however, it is thought safer to isolate the speaker windings by a condenser in order that users may not receive electric shocks from the terminals or speaker leads, if for no other reason.

The second diagram of Fig. 126 (*b*) shows a transformer coupling from anode circuit to speaker. This has similar advantages, isolating the speaker, providing by a primary winding of thick wire a low-resistance anode circuit, and at the same time matching anode to speaker impedance when these are unequal. We shall see that this is necessary in most modern types of speaker, and that it is the most widely used arrangement to-day.

If the turns ratio of such a transformer be T, then a speaker or other device of impedance Z_s will be equivalent to an impedance Z_s/T^2 in the primary circuit. In this way, therefore, any speaker, particularly if its impedance be much less than R, can be matched to the anode circuit and the maximum power output developed in it.

Up to the present we have assumed that the load resistance R is independent of frequency, and therefore that the same power output is available at all modulation frequencies. This is fairly true of a resistive load except at very high frequencies, when stray shunt capacities may play a part, but it is not necessarily the case of an inductive load, such as a speaker, choke, or transformer primary. Taking the simplest case, of a choke, the inductive reactance will fall with frequency, becoming zero at zero frequency, when only the ohmic-resistance residue remains as anode load. Hence, if the inductive reactance were chosen to equal the optimum load R at a very high frequency, the efficiency of the stage would fall uniformly to zero at the lowest frequencies. In practice, when a choke is used to complete the anode circuit it is chosen to have a very high inductance, of from 20 to 100 henries, and thus greatly exceeds R over most of the frequency range, leaving the lower-impedance speaker to

act as the determining factor. At low frequencies, however, the choke impedance falls to the same order as R, or even lower. The total impedance falls progressively below R, and the power output drops off also. It is thus most difficult to maintain uniform efficiency below some 50 or 100 cycles.

Since it is the speaker which eventually furnishes the useful part of the load in any form of output circuit, it is desirable that its impedance should be uniform with frequency. With the majority of speakers this is seldom the case, the tendency being for a mainly inductive impedance rising with frequency and showing a variety of irregular peaks and troughs at frequencies where mechanical resonances occur. Reproduction cannot therefore be uniform. The tendency in speaker design is of course to eliminate such changes in impedance as far as is possible. They must also be considered in relation to the efficiency with which the speaker converts electrical into sound energy. This also generally rises with frequency up to some high limit, being poorest in the lower register, and so the best procedure is to match the speaker impedance to the power stage for maximum efficiency at a low frequency, probably between 50 and 250 cycles per second. The reproduction will then fall off quickly below that figure, but above it the rising speaker efficiency may largely compensate for increasingly poor impedance matching, and a fairly uniform over-all efficiency result. At some figure between 4,000 and 10,000 cycles, according to the type of speaker in use, the efficiency of conversion into sound will cease to rise. Since the impedance matching will be very poor also at so high a frequency, a virtual 'cut-off' takes place. Thus it is clear that the reproduction from most speakers will be fairly uniform over the main audible range but cease almost entirely below and above certain limits.

It has been pointed out that the efficiency with which a triode power stage converts anode-current input into electrical-power output lies between 20 and 25 per cent. only. This implies unnecessary expense in providing anode current. In the pentode, however, a more efficient converter exists. Fig. 127 shows a family of anode current-anode voltage curves similar to those for the triode in Fig. 125. It will be noticed that the pentode tends to operate as a constant-current device. The curves show that, once the very curved portion at low anode voltage has been passed, the working portion beyond A indicates only a

small change of anode current with increasing anode potential. Under these conditions of comparatively constant current change, the maximum power will be developed when the load impedance is very high, increasing in proportion to impedance over a wide range. The two example load lines shown illustrate this point, since the area of the triangle *ACB* increases as the

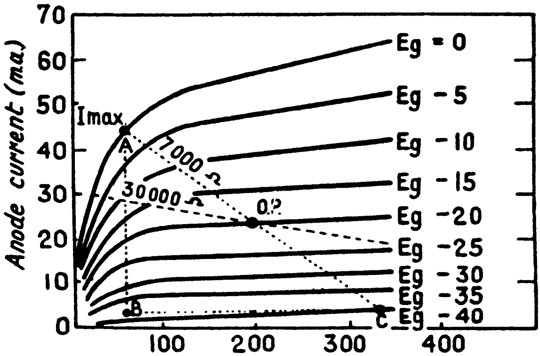


FIG. 127.

load line becomes more nearly horizontal, corresponding to values of load impedance tending to infinity.

On account of this approach to constant current, the pentode when used with an inductive load must inherently give rise to distortion, the efficiency increasing indefinitely as frequency or load impedance rises. Thus, since speaker efficiency tends also to rise with frequency, the higher frequencies will be progressively better amplified, and reproduction will tend to sound very 'top-heavy'. This is sometimes convenient, as when we wish to bring up the high-frequency response after it has been reduced through excessive selectivity. It must not be overdone, however. For one reason, very high oscillatory potentials may be set up across the load at the highest frequencies where current will be low and power output a maximum. These are capable of damaging the equipment. Also, excessive response above the main audible range will accentuate harmonic distortion and will accentuate any valve hiss or background noise.

Fortunately a simple filter will compensate for this rising characteristic. A condenser in series with a resistance is shunted across the load, the resistance being low enough to limit the total impedance of the two in parallel to a safe value. At low

frequencies the condenser will possess a very high reactance, thus virtually removing the resistance from circuit. As we ascend the frequency scale, the condenser reactance falls, and an increasing proportion of the oscillatory current will flow through the resistance, instead of increasing the output of the speaker. At very high frequencies the condenser reactance can be neglected, and the resistance is virtually in parallel with the speaker. By suitable choice of resistance and condenser, the characteristic can be altered as desired or can be made as level as that of a triode stage. Thus it is safe to say that the quality obtainable from a pentode is equal to that of the triode, provided too great an output is not attempted.

The characteristic of a pentode can be shown to result in a larger percentage of harmonic distortion when anything approaching the full power output is made use of, the curvature of the dynamic curves being normally greater. The third harmonic is the most strongly produced, and gives rise to a slightly harsh tone-quality characteristic of a fully driven pentode. This is sometimes termed 'pentode tone', and cannot be fully removed by a filter without at the same time weakening the high-frequency reproduction. It is considerably reduced when a pair of pentodes are used in push-pull. The use of a pentode is apt to accentuate any resonant peaks in the loud-speaker response, because at such points speaker impedance rises. Therefore the pentode delivers more energy and makes the peaks more noticeable.

To compensate for slightly greater harmonic distortion, the pentode offers several other advantages, which can be appreciated from the curves. The grid swing required to change the anode voltage through wide limits will be seen to be much less than for a triode of similar output. The amplification factor is higher, and in consequence less signal amplitude is needed to drive the pentode. When first designed, the valve was mainly regarded as a means of cutting out one amplifying stage, but to-day it is more valuable on account of its better power-output efficiency. This is due to the fact that power output rises with load impedance, and hence with frequency. Thus if we make use of a higher load, matching our impedances for a somewhat higher frequency than for the triode, a better average efficiency is to be expected. This is quite a practicable procedure, but it is difficult to state figures because the efficiency varies so widely

with design. Fifty per cent. efficiency can be obtained from a pentode stage, but with somewhat increased harmonic distortion. When economy is important, however, the pentode is most valuable, since it reduces both the anode current and voltage necessary for a given acoustic power output, and also the amplification which must be used in preceding stages. It may be quite possible to feed a pentode directly from the detector, a diode being able to deliver from 10 to 30 volts of signal amplitude with ease. This is sufficient to drive a large pentode, but would not be sufficient for the corresponding triode.

Single-pentode or triode output stages are satisfactory when no enormous power output is desired, the former being seldom used to deliver more than 10 watts of electrical energy, whilst the latter may be used at larger outputs than this in special cases demanding an absolute minimum of distortion. It leads to very high cost, however, since the anode current and voltage must be large, and the components to carry these expensive and bulky. The use of two or more valves in parallel to increase output is perfectly possible, and was common at one time, but it has now given way to the use of push-pull. An extraordinary increase in efficiency is found possible by this method, and certain recent developments from it have made the single-valve output stage obsolescent at higher powers.

The use of two valves in simple push-pull is shown in Fig. 86, the basic circuit being already familiar. If each valve is biased to the usual class A position, the output will be approximately twice that of each singly, as would be expected. This is a better way of using a pair of valves to do the work of one larger one, however, because we have seen that push-pull results in cancellation of the even harmonics introduced by the effects of curvature. It is thus possible to improve quality, or alternatively to operate each valve over a longer portion of its characteristic, with rather more than twice the output of each singly. Greater grid swing is of course necessary to do this, but there is still little input power necessary, since no grid current should flow. For best quality the valves should be matched, and it may be convenient to arrange for individual grid biasing as shown in the diagram, so that each anode current can be adjusted to equality. We have already noted two other advantages of push-pull operation. The first is that the balanced output transformer prevents unwanted coupling to or from the high-

tension supply, hum being largely cancelled out; the second a reduction in the bulk and cost of this transformer brought about because the two primary currents magnetize the core in opposition, the mean magnetic flux being zero. A small iron section can now be used, sufficient to handle the alternating components of flux only. At high powers this saving can be very considerable.

The above facts alone make a good case for the adoption of push-pull, but they are not the most important. In class A working each grid receives the whole signal wave form, the two valves amplifying normally but in opposite phase. Consider what will happen if a larger bias be employed, such that each valve is worked at the class B condition defined in Chapter VIII. The negative bias is such that each valve works near the bottom bend of its characteristic, almost no anode current flowing in the absence of a signal at the grid. Now the alternating potentials to be amplified are applied in opposite phase to each grid, and so, whilst the positive original half-waves will be amplified to produce anode-current pulses in the valve A, the originally negative half-waves will be reversed in phase by the input transformer and applied as positive half-waves to the grid of B, being in turn amplified by that valve only. Since each valve is negatively biased to anode-current cut-off, no further reduction is possible, and negative potentials added to the grid bias cannot produce energy in the output circuit. Thus each valve now amplifies half of the input wave form only, the other valve amplifying the second half.

This sharing of the input between two valves so that each amplifies alternate half-cycles has many advantages. In America it is regarded as a particular case of class B audio-amplification, whilst in Great Britain it has been given the special title of 'quiescent push-pull', the word 'quiescent' implying that the total anode current is practically zero in the absence of signals. At once a valuable property becomes evident, for the absence of anode current when there is no modulation will be a decided economy in high-tension consumption, invaluable in battery-operated receivers. It should not be difficult to see that the anode current of this type of power-output stage will rise in approximate proportion to the input potentials, the mean anode current varying as the 'grid swing' or mean amplitude of modulation. Hence the anode current will be self-adjusting to a value just sufficient to handle the signal strength at any moment.

This wide variation of mean current must not be accompanied by any corresponding changes in mean anode voltage, however, or distortion will result. The high-tension supply must be closely independent of load, and must have a low internal resistance. A good battery or accumulator meets this requirement, but it is less easily met by battery-eliminators, in which the chokes must be of exceptionally low resistance, the filter of the inductive-input type, and the rectifier of low internal resistance such as a mercury-vapour diode. Similar steps must be taken to see that grid bias also remains independent of anode current.

Biasing Methods. At this juncture we can conveniently digress to consider the methods by which the grid bias for amplifiers is obtained in practice, and how it can be kept uniform when necessary. In the chapters on transmission it was pointed out that grid bias for an oscillator can be obtained simply from the potential drop produced by grid current flowing through a grid leak. As a rule this method is quite useless in amplifier circuits, because in the most usual class A condition the grids must always remain negative and no grid current flows. We shall see in a moment that under class B power-amplifier conditions some grid current is allowed to flow, but in this special case it will vary widely in magnitude with the amplitude of the input signals. It is thus unsuited to grid-bias purposes when the amplifier is to work with little distortion over any wide range of frequency or signal strength. The general practice in this case is therefore to keep the grid-circuit resistance very low when there is any possibility of grid current, so that no appreciable bias can arise from this cause. In certain cases of modern high-power low-frequency amplification, grid-current bias is appreciable and must be allowed for, but such cases are beyond the scope of this book.

In all early circuits grid bias was obtained from a battery and was thus constant and easily adjusted. Battery bias is still very satisfactory, for if the battery be reasonably fresh and in good condition its voltage can be taken as constant for all practical purposes, whilst its internal resistance will be negligible. Battery bias is therefore used in most receivers which also employ battery high- and low-tension supplies, in laboratory work, and in many other uses of amplification. It needs little comment, and cannot very well give rise to distortion or other defects,

provided that the battery condition is closely checked. The use of direct-current generators as a source of bias in transmitters or very large power-amplifiers has already been noted. Separate battery-eliminators similar to those used for anode-current supply are also suited to biasing work when convenient. Their design is made very simple owing to the minute current needed,

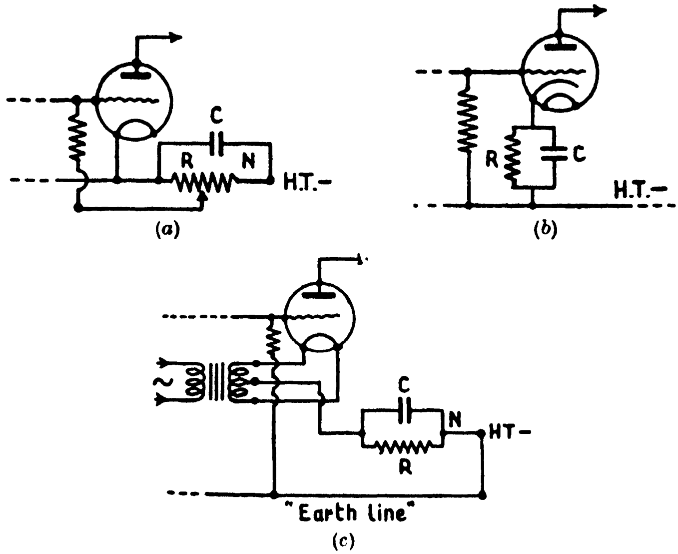


FIG. 128. Grid-biasing Circuits.

allowing high-resistance chokes wound with fine wire, and needing low values of smoothing capacity. A resistance will preferably be joined across the output from a grid-bias eliminator, taking a load of a few milliamperes. This serves to stabilize the voltage output, and may be tapped or treated as a potentiometer to provide one or more lower voltages.

Two very convenient methods of biasing exist which are applicable to the case of mains-operated receivers and amplifiers in which indirectly heated valves are used. The first of these is illustrated in Fig. 128 (*a*) and the second in Fig. 128 (*b*).

In the first system a resistance is inserted in the negative lead from the receiver to the high-tension supply, all the current drawn by all valves thus passing through it. The value of this resistance is so chosen that the highest grid-bias voltage needed will be built up across it, and the negative end of the resistance N is treated as the negative biasing source. Lower voltages can

again be tapped off at intermediate points along this resistance, as shown.

The second system is termed 'cathode bias', and is by far the most widely used to-day. Here a resistance is inserted in the cathode lead of each valve. The mean anode current of the valve flows through this, producing a potential drop across it which biases the cathode positively relative to the 'earth line' or negative high tension. The grid return circuit is joined to this earth line, however, and hence the cathode is biased positively with respect to grid, which is equivalent to a negative biasing of the grid with respect to cathode. The biasing resistance R is chosen from Ohm's Law so that the potential built up across it by the correct mean anode current (plus screen current in the case of screened-grid or pentode valves) will set up the correct grid bias wanted. For example, if a valve is to work at a point upon its curve at which the anode current is 10 milliamperes and the corresponding negative bias 10 volts, a cathode resistance of 1,000 ohms will provide this.

Cathode biasing is cheap and convenient. It has the advantage that each valve can be separately biased to the best advantage without reference to others, whilst in the previous method, employing a single resistance in the negative high-tension lead, the anode current of each valve will affect the bias of all others. Both methods, and particularly the cathode system, have the advantage of being partially self-compensating or automatic. If, for example, the anode voltage of a valve falls, as when a high-tension battery runs down, the anode current will tend to drop also. This will reduce the current through R, and hence the grid bias in an equal proportion. A constant ratio of bias to other operating potentials and currents is thus set up, which tends to maintain the valve near the correct working-point on its characteristic in spite of quite large variations in other factors. This condition makes for ease of design and adjustment, and for reliability. It also reduces the tendency to overload in amplifiers: for suppose an excessive signal results in an unusually large peak of anode current; this will set up a correspondingly large momentary grid bias which operates to prevent the grid from becoming positive under any conditions. As a result, a power-amplifier will often deliver somewhat larger output for a given degree of distortion if it be biased by one of these semi-automatic systems.

It is not thought necessary to complicate the diagrams in other chapters by showing practical grid-biasing circuits in each case. Very often battery bias is shown or implied when practically one of the methods just described would be used. This should be borne in mind, together with the fact that in perhaps nine cases out of ten each valve of a modern receiver will be biased correctly by the cathode system of Fig. 128 (*b*). This can also be applied to directly heated stages, as shown in Fig. 128 (*c*), where a centre tap on the filament circuit or transformer is taken to negative high tension through the necessary biasing resistance. Observe also that these methods imply a constant *average* anode current, which will be found under class A conditions. They may not be suitable to systems in which normal current is nearly zero or varies widely with signal amplitude, as in class B stages, in which a separate biasing source is the safest method to select.

The condenser C in each diagram is most important, as it serves to smooth out the rapid fluctuations of bias potential which would otherwise take place at signal frequencies. C and R together form a filter network of the simple type discussed in the chapter on detection. The alternating component of anode current flows through the low reactance of C, whilst only the mean current provides bias across R. It is essential for C to be large enough for this to take place, if the cathode-biasing system is to behave correctly.

C has another important function. It will be noticed that the cathode resistance is common to both the anode and the grid circuits of the biased valve, and we have seen that an impedance common to two circuits forms a coupling between them. It may therefore throw the stage into oscillation or seriously upset its working. To overcome this the impedance of the cathode circuit must be made negligible at signal frequencies, which is the responsibility of the condenser C. This must have a reactance which is negligible relative to that of the anode or grid circuits. Suitable capacities will be of the order of 0·1 microfarad at radio-frequencies, but will need to be several microfarads in low-frequency amplifiers.

In the latter case C can form a source of distortion, since as zero frequency is approached its reactance cannot remain small. Reaction therefore creeps in, and since anode and grid potentials are normally in opposite phase this will be negative reaction

and will reduce the amplification of the stage at low frequencies. It is in fact one of the chief sources of loss in that region. If C be increased to at least 50 microfarads, negative reaction is not serious down to some 20 cycles per second. Special low-voltage electrolytic condensers are used in this position, of the largest capacity that cost will permit. No matter what grid-biasing system is used, it must have a very low impedance to the lowest frequency desired, and will often be shunted by a large condenser to ensure this. Subject to that qualification, grid bias need not be discussed in relation to each type of amplifying circuit studied, it being taken for granted that a suitable system has been selected, capable of maintaining the bias voltage steadily during normal conditions of input or output amplitude. We will now look a little further into class B power amplification on that assumption.

Class B Power-Amplifiers. We have seen when discussing amplifying stages in transmission that the power-conversion efficiency rises as bias is increased to the class B condition. This applies also in the present case, the ratio of anode input power to signal output power improving. Thus a quiescent push-pull amplifier will deliver more power to the loud speaker for a given high-tension consumption; but remember that each grid can now accept twice the potential swing, and that the input amplitude will need to be doubled for this improvement in efficiency to be realized in practice.

The distortion of a quiescent power-amplifier would seem at first sight to be serious, and is in fact somewhat greater than for a class A biased triode. Part of the grid swing will occur over the curved bottom bend of the characteristic, and harmonics will be introduced on that account; but fortunately these will be identical for each valve and will largely cancel out in the differential-output transformer. The design of the latter is complicated by the fact that the impedance of valves used in this manner will not be independent of amplitude, but will fall as the anode current rises during strong signal periods. Fortunately the natural tendency is for transformer primary reactance to fall also with increasing current, and it is found possible to design transformers which will minimize both these defects to a large extent. The transformer core must be large enough, however, to prevent saturation during loud passages, when both the direct-current and signal components are at

their maximum. The average alternating-current resistance of the pair of valves looked at from the secondary winding of this transformer will be double that of a single valve, rather than half, as in the case of valves in parallel, and this fact must be taken into account when matching to the speaker impedance.

The quiescent push-pull power stage offers numerous advantages, particularly when of small power output and fed from a limited supply. Ordinary triode valves may be used and, as in transmission, may with advantage have somewhat higher amplification factor and impedance than usual output triodes. Pentodes can be used in push-pull or class B quite successfully. Special valves are also made in which two matched triodes of appropriate characteristics are enclosed in a common envelope, and it then becomes possible to employ a close-mesh grid, so that the operating-point comes near to the bottom bend with little or no added grid bias.

In all the cases so far mentioned the grids are driven from a high negative value up to zero grid voltage, no grid current being allowed to flow. Whilst therefore twice the grid swing needed for a single valve must be provided by the preceding stage, there is no considerable power absorbed, and this stage can be designed purely as a voltage-amplifier.

Still further power output and efficiency should be obtained from the same valves if the grids are driven positively, so that the whole length of the characteristic up to anode-current saturation can be used instead of only that portion on the negative side of zero grid potential. This is strictly comparable to class B or C working in transmission, and we have seen that it improves efficiency still further, but necessitates a large input to the grids backed by power sufficient to provide the grid current which will now flow. The term 'class B power-output stage' is now universally used to denote such conditions, with perhaps the qualifying description of 'positive drive' in addition.

A positive-drive class B stage is designed on similar lines to those already discussed, and on the anode side will be exactly similar, with a larger output transformer to handle the larger peak currents, and matched to the somewhat lower impedance which the valves acquire when driven positively. On the grid side, however, special precautions are necessary. The preceding stage must be designed as a small power-amplifier, and it is

usually necessary for it to deliver about the same signal power as would the final valves if used in class A. This is now termed a 'driver' stage, and can best be a class A triode, when the combination of driver and class B power stage is often treated as virtually a single stage. The driver will thus be a valve of the same type or similar rating to those used in the push-pull position. Its efficiency has been shown to be some 20 or 25 per cent., whilst that of the class B stage may reach 70 or 80 per cent. when driven positively, giving an output from the final stage of about ten times the signal-power output of the driver, and for only about three times the mean input power to the three anodes.

For distortion to remain low under positive-drive conditions it is necessary that the grid wave form remain unchanged both when negative and when positive. In the former case the grids represent a very high impedance, whilst as soon as they become positive they represent a comparatively low resistance which decreases with potential applied. To preserve wave form under this very asymmetrical loading power must be supplied and the resistance of the grid driving circuit must be low. Little potential drop will then occur in the transformer secondary for large values of peak grid current. A transformer is usually employed, as in previous push-pull circuits shown, but instead of being a step-up transformer having high secondary resistance, it will now require a small step-down ratio and a secondary resistance not exceeding a few hundred ohms.

This transformer is designed similarly to one driving a loud speaker or other load. The impedance of the driver anode circuit is known, and the mean impedance represented by the class B grids when positive can be estimated from the grid current-grid voltage curves of the valves used. The transformer ratio is chosen to match these two, as previously explained, and since the grid impedance is usually the lower the ratio may be a downward one of some 2:1. Since the driver will have a low amplification factor also, there will be little or no gain in this stage, its purpose being solely that of providing the essential driving energy for the main power stage. The latter also has little voltage amplification, and it is quite possible that this factor may be unity or even less for the whole output system, the potential applied to the driver grid being sometimes as large as that delivered by the output transformer. A very large

power amplification occurs, however, reaching several hundred to one in a well designed circuit.

Positive-driven class B stages lend themselves to the generation of really large outputs, being used mainly when the simpler forms would be uneconomic. Less than some 10 watts of signal power can be effectively produced by class A methods, but for powers exceeding this and up to many hundreds of watts the class B system will give an enormous saving in cost. This occurs partly on account of the smaller valves that can be used, the low values of mean anode current, which reduce transformer costs, and the economy in total high-tension current consumption. By working the power valves over the whole of their characteristics a large gain in output from any given valve is obvious. The peak current obtained is limited only by valve-cathode emission, special types being advisable for high-power work, in which this factor is specially copious. The ratio of peak current to mean current is very large, the latter being low, because zero current flows during at least half of the cycle, and efficiency has been shown to depend upon this. The value actually obtained depends upon very many design factors, but may be five or six times greater than for class A. As an example, two valves capable of an output of 6 watts of signal energy in class A may deliver 60 watts in class B when fully driven. Such outputs are not often necessary in radio-reception, but are very useful in public-address work, or in the modulation of transmitters.

This chapter commenced by the consideration of faithful reproduction in radio-reception, which led to an analysis of the power-output stage, in which much of this can be lost. The position in this respect can be summarized by stating that a large triode used in class A is likely to be the best when distortion is the main consideration. The total harmonic content introduced by this stage need not exceed 1 per cent., varying from about 2·5 per cent. under full load to less than 1 per cent. under ideal conditions. When larger power output is needed with economy, the pentode or simple push-pull may be resorted to. The harmonic content will remain small at moderate output, but may rise to perhaps 5 or 7·5 per cent. when the full output is being taken. Still larger output with greater economy demands positive-driven class B stages, the distortion of which depends very much upon skill in design but may easily exceed these

figures at maximum output. In all cases, however, the simple expedient of working the output stage at least 50 per cent. below its full output capabilities will greatly improve harmonic distortion, reducing it to negligible proportions in many cases. It can then be said that the wave form delivered to the loud speaker is virtually that entering the power-output stage.

The Loud Speaker. Having produced sufficient signal power at a low level of distortion it must be made use of in some recording or indicating device, the only one to be considered here being the loud speaker. In Chapter VII the construction of the simple telephone ear-piece was described. For many years this remained the only generally used device for the transformation of electrical energy into sound, and it is still extensively found in very much its original form. An improved type used during and after the War was manufactured by S. G. Brown, Ltd., and known as 'A' type. Here a steel reed vibrated across the poles of an efficient magnetic system, the air gap being very small and the magnetic efficiency high. To this reed was fixed a light aluminium cone which replaced the diaphragm of earlier telephones and set the air into vibration. The sensitivity of these telephones was an improvement, and they are still popular with commercial operators.

Another striking step forward is the introduction very recently of piezo-electric telephones, in which the vibrations of a Rochelle-salt crystal set up the necessary sound-waves. The low inertia of such a crystal and its sensitivity to high frequencies, mentioned when describing microphones, result in a telephone which will reproduce very much better over the upper audible range than its predecessors. Moving-coil telephones are said to be under development at the present time, in response to a demand for high quality that has grown up recently in the monitoring of sound-recording equipment.

Whilst telephones remain the favourite instrument of the professional telegraphist, partly on account of their effect in keeping out external noises, the loud speaker is used almost exclusively for broadcast and general reproduction of telephony. Early loud speakers consisted simply of large telephone units fitted with more or less efficient horns to amplify the sounds they produced. Their reproduction was therefore characterized by poorness of high notes and complete absence of frequencies below about 200 cycles, which effects were typical of magnetic

telephones, added to a certain amount of resonance from the small horns usually employed. There is nothing against the horn-type speaker, which to-day is widely used in powerful equipment; but it is necessary that the horn be large if it is to amplify low as well as high notes and not to resonate within the main acoustical range. Early types were far too small to meet this requirement and reproduced only over a frequency range of perhaps from 200 to 3,000 cycles.

Gradual improvement took place in the design of magnetic units to eliminate resonances and extend the range. The steel-reed type proved the better, being able to vibrate at larger amplitudes than a flat diaphragm, without so great a distortion of wave form. Such reed units were attached to many forms of large slack diaphragm capable of setting the air into vibration without the help of a horn.

In common with all other power transforming devices, the speaker calls for correct matching of the input and output impedances if efficiency is to be high. The input impedance is of course that of the electrical circuit, modified considerably by the reaction upon it of mechanically vibrating parts such as the reed. The output impedance is an acoustical one, represented by the loading imposed upon the 'diaphragm' by the outside air. Air is a viscous medium which will oppose any object vibrating within it, energy being absorbed from that object and radiated as sound-waves. It is typical of such a reaction that it will be greatest at high rates of vibration, falling at lower frequencies, to become zero at zero frequency. A speaker will deliver power most effectively to the air when the acoustical impedance is correctly matched, and for a uniform sound-output at all frequencies this loading should be kept at the optimum value. A compromise at once appears, for, whilst a constant loading of the diaphragm is necessary for uniform efficiency, that actually imposed by the air falls off with frequency.

One method for overcoming this lies in the use of a horn, the air column in the neck of which imposes a load upon a diaphragm vibrating within it. This loading is of the same order as that imposed by the unconfined air on a diaphragm equal in area to the mouth of the horn, and so the addition of a horn is equivalent to the use of a very much larger diaphragm. A suitably selected horn, particularly if exponential in shape, will

load a diaphragm amply and with little variation over the whole range of frequencies, from the highest at which it can vibrate down to a lower limit. This limit occurs when the length of the horn equals rather less than a half-wavelength, and below it the loading quickly disappears. Thus a large and long horn is needed for efficient reproduction of the lower notes, a length of 8 feet being considered a minimum. A 4-foot horn may suffice for speech, when a lower limit of some 200 cycles will suffice. Above this cut-off frequency, however, the loading of a horn is excellent and the efficiency of the speaker to which it is attached will be as high as its electrical design allows.

The loading of a diaphragm by free air naturally increases in proportion to the area of the active surface. The alternative to the use of a horn is thus a large diaphragm, and the alternative development of speakers has been in that direction. The vibrating steel-reed unit was first attached to various forms of 'diaphragm' free to follow the movements of the reed as far as possible. Pleated paper, stretched metal, and wooden disks are amongst the materials tried, but a stiff cardboard cone was found to be the most generally suitable object, and the popular 'reed-driven cone speaker' still used in inexpensive receivers came into existence. Theoretically the cone was expected to vibrate as one unit, with a piston-like motion. Practically, however, research showed that it is impossible to arrive at a combination of lightness and rigidity which will achieve this. At very low rates of vibration the whole cone may move, but at some frequency well below 100 cycles the cone breaks up into more complex modes, vibrating in numerous sections. Since these do not move in phase, their effect upon the air is not additive, and the loading varies irregularly about some mean value. Resonances within the structure of the cone invariably occur, causing certain frequencies to be favoured whilst others are poorly reproduced, and a main cone resonance usually occurs somewhere between 1,000 and 3,000 cycles. Above this only the central portions of the cone vibrate strongly, and the loading and efficiency fall steadily.

Since loading by the air falls with frequency, it is only possible to maintain reproduction if the diaphragm can be made very large, or its amplitude of vibration increased proportionately. For frequencies below, say, 100 cycles per second, such as are contained in the beating of a drum, the necessary amplitude

may reach a considerable fraction of an inch. A steel reed moving very close to pole pieces cannot reach an amplitude of this order, and so reproduction in the extreme bass is impossible to speakers of this variety. The use of a horn in conjunction with a much smaller cone is clearly superior, since the amplitude required is not so great; but a very large horn is essential if the lower musical frequencies are to be reached. These are taken as from 20 to 50 cycles by different authorities, and necessitate a horn upwards of 30 feet in length. Such horns can be used with conspicuous success in cinemas or large installations but are domestically inconvenient. The ordinary cone speaker or early horn type thus reproduces very little true bass, has a very irregular response curve at medium frequencies, and falls off seriously towards the upper limit. It is thus decidedly the weakest link in radio-reproduction, and we must thank the surprising accommodation of the ear that the reproduction furnished by it was acceptable for so many years.

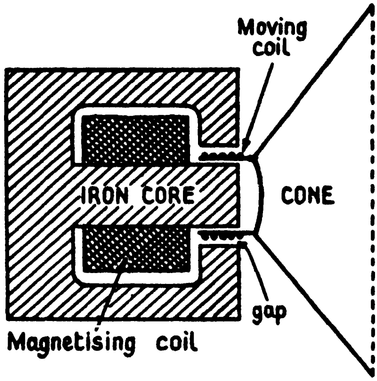


FIG. 129. Construction of Moving-coil Speaker.

A big advance in true reproduction came with the development of the moving-coil speaker, with which the names of Rice and Kellog are usually coupled. In this design the large cone is driven from its apex by a light coil of wire free to move within the field of a concentrated magnetic system. Fig. 129 illustrates the general construction of such speakers and of most other moving-coil devices such as microphones. The magnet comprises a central pole of circular section, round which is wound the exciting coil if the magnet is not a permanent one, and the other pole terminates in a ring-shaped structure leaving a small circular gap between the poles. The thin cylindrical coil attached to the cone lies in this gap without touching either pole, being maintained at a few thousandths of an inch from these by 'spiders' or flexible supports. These allow the freest possible motion of the coil along its own axis when the turns cut across the lines of magnetic flux crossing the annular

gap. Hence the coil will experience forces proportional to the currents flowing through it which will set the attached cone into vibration at the same frequency as these currents. The magnitude of these forces will be proportional to the current through the coil multiplied by the magnetic-flux density within the gap, which should therefore be as high as possible. This varies from 10,000 up to nearly 20,000 gauss for modern magnets, the figure being somewhat better as a rule when an electromagnetic system is used. The consumption of this system may lie between 10 and 30 watts, and can be derived from the supply mains, any convenient direct-current source, or sometimes from the eliminator circuits of a receiver, in which this electromagnet may be used as smoothing choke to assist the filter system.

The moving-coil-driven cone will vibrate against a mechanical restraining force composed of the resistance of the centring supports, the inertia of the cone and coil, and the air damping. At low frequencies most of these become small, and the free suspension will allow a large amplitude of vibration to build up, resulting in a much improved bass response. At medium frequencies the resonance effects due to a stiff reed are eliminated, whilst reduced mass may improve the response near the upper limit in many cases. Cone resonances still remain, but on the whole there will be a more uniform characteristic, better maintained at the extremes of frequency.

The small size and essential lightness of the coil, however, make it most difficult for this impedance to be high, 'high-resistance' coils being possible but unsatisfactory. It is more usual to build the coil of a few turns of fine wire, it having a resistance of between 20 and 2 ohms and an impedance averaging about four times this figure. Coils exist comprising a single turn of aluminium strip only. Hence an impedance-matching transformer is necessary, the ratio of this being of the order of 25:1, which we have seen fits in admirably with the design of modern power-output stages. Impedance should be matched at a low frequency, at which response has commenced to fall off, and this helps to maintain an even response.

A moving-coil speaker used in the open will be found to reproduce very little at low frequencies, even if the cone be vibrating freely. This is due to interaction between sound-waves produced from the front and back of the cone. For low acoustical frequencies having a long wavelength the front and back

of the cone are nearly at a single point in space. Sound-waves radiated from each must be in opposite phase, since the back of the cone will be receding when the front is advancing, and vice versa. Hence the conditions are favourable for complete interference between the two sets of waves. At higher frequencies, where the wavelength becomes a few inches only, the front and back of the cone will behave as independent sources, not necessarily in phase opposition, and elaborate interference patterns will surround the speaker.

To overcome all these troubles a 'baffle' is used, consisting of a large sheet of non-resonant material which should be a poor conductor of sound. The speaker is mounted through a hole in this baffle, which serves to separate sound-waves produced by the front and back of the cone, and to check interference between them. The efficiency of the speaker is then rendered more uniform with frequency, and the lower tones will be fully reproduced down to a rather indefinite limit occurring when the width of the baffle surface becomes of the order of a half-wavelength. Below this limit interference can still operate to reduce the output, but if the baffle area exceeds at least 30 square feet reproduction will not fall off seriously above 50 cycles.

The cabinet in which a speaker is mounted often provides a baffle, and may be adequate if large and strongly built of thick material. The presence of an enclosed volume of air behind the speaker is undesirable, however, as it may resonate within the audible range, producing a 'boomy' or 'boxy' tone. Baffle plates may be added within the cabinet, which will reduce the tendency to resonance by breaking up the enclosed air into smaller unequal sections. Absorbent material is used to line the inside of speaker cabinets, thereby reducing reflection by the walls, and damping out resonances. It is also possible to terminate the rear of the speaker by a scientifically designed 'acoustical impedance' or 'labyrinth' in which sound will be completely absorbed, a method adopted in a few American receivers.

The placing of a speaker within an aperture in an opaque baffle or wall increases the loading of the diaphragm, and it is thus a substitute for a horn. These two methods of loading are employed alternatively to suit particular conditions, it being possible to attain the required acoustical load either by the use of a small diaphragm fitted with a large horn, or by a larger dia-

phragm (or cone) fitted to a large baffle. The characteristics of the two systems differ both in frequency response and directional properties, the latter being often a deciding factor in practice. The efficiency at very high frequencies differs little, depending mainly upon the electrical design of the speaker. Over the medium-frequency range this also applies, but in addition the better loading obtainable from a large horn results in a better ratio of sound-energy to electrical input. In the lower register it is found that the horn-loaded speaker maintains its output best until a limit depending upon horn length is reached, when the reproduction practically ceases; whilst the open-baffle speaker falls off steadily with frequency as zero is approached. The cut-off is less sharp in this case, and hence the extreme low notes may be better maintained.

The directional effects of an open baffle are not very pronounced, although there is a tendency for the higher frequencies to be 'focused' into a beam along the axis of the speaker cone, the width of this beam becoming less as frequency rises. The horn speaker, on the other hand, is directional at all frequencies, projecting most of the sound-energy into an area bounded by an imaginary continuation of the horn itself. In an enclosed space reverberation will soon mask this effect, but in the open air a directional characteristic is useful in concentrating sound where it is wanted. For all classes of public-address work, in its widest sense, the horn-loaded types are preferred, both for the above reasons and because of their better average efficiency.

It is very difficult to design moving-coil loud speakers which will reproduce equally well at the two extremes of frequency, since the requirements of each case are in many respects opposed. For the best reproduction possible it has become usual, therefore, to combine two or more speakers, each speaker of the group favouring a part of the musical scale. The simplest case consists of a 'matched pair' of speakers, one of which will employ a large cone, giving good efficiency at low and medium frequencies. This cone will possess too much inertia to respond well at the highest frequencies wanted, and so the second speaker will employ a light rigid cone, designed to reproduce most effectively above the point at which the former commences to fall off. This high-frequency speaker has been termed a 'tweeter' and can with advantage employ a small horn, a foot or so in length. An electrical filter network may be provided to ensure that the

most suitable frequency range reaches each speaker, a slight overlap being provided in the middle register. In the design of large loud-speaking installations it is becoming usual to provide two groups of speakers, handling the upper and lower portions of the spectrum in this manner.

From time to time other electrical effects have been pressed into service for loud-speaker design, generally without striking success. In fact, all those principles of operation described under microphone construction can be used with varying success in speakers also. We have described the uses of magnetic attraction, firstly between a fixed-magnet system and vibrating armature, as in the telephone earpiece and the reed-driven cone, and then between a moving coil and fixed magnetic field. Electrostatic forces have been similarly used to produce 'condenser' loud speakers, on the general principle of a massive fixed back-plate and a stretched conducting diaphragm parallel and close to it. Alternatively both electrodes can be slack and flexible, such as metal-foil sheets separated by corrugated paper. The amplified signals are applied between these two electrodes, which form the plates of a condenser and are held in a state of tension by a fixed potential between them. The alternating potentials add and subtract from this fixed polarizing potential, varying the attraction between the two electrodes and setting one or both into vibration. The force between two such charged bodies varies directly as the potential difference and as the area, and inversely as the square of the distance between them. For large forces to be built up a large area of surface is necessary, which is desirable in loud-speaker design since it increases the acoustical loading.

A very small separation is also necessary, and is a limiting factor, for the amplitude of vibration of the moving electrode cannot exceed this. The electrostatic speaker cannot reproduce low frequencies effectively for this reason, and has never become satisfactory for general use. At high frequencies, where a very small amplitude suffices, it is possible to decrease the electrode spacing, thus increasing the attractive forces considerably and reaching a high order of efficiency. Difficulties of insulation at the high voltages necessary have been troublesome, but several excellent speakers of this type have made their appearance in the 'power-tweeter' class, that of Siemens in Germany being a prominent example.

The piezo-electric phenomena has been applied to speaker construction, but with less success than to microphones or telephones. Very good high-frequency response is possible, combined with good sensitivity, but the crystals cannot oscillate with sufficient amplitude to provide large output in the bass. They are also somewhat fragile. The freedom of movement essential for bass reproduction seems only practicable on the moving-coil system, and at the present day the development of other types seems largely in abeyance. Moving-coil speakers are being slowly but steadily improved by continual attention to detail and by progressive elimination of all resonances objectionable to the ear.

Fortunately the human ear seems unaffected by the phase relationships between the components of a complex wave, and so it is not important that these may be seriously modified in the electrical circuits and the speaker. The acoustics of any ordinary room would in themselves be sufficient to change the phase relationship of a mixture of frequencies, and the fact that music from a distant band remains undistorted shows that in this respect phase plays no part. It is established that the smallest change in volume noticeable by the ear is about 2 db., a change of 5 db. being generally regarded as slight. The modern speaker will reproduce with less variation than this a frequency spectrum of from 200 to perhaps 5,000 cycles, which we have seen to be the main audible range. Below 200 cycles most speakers mounted upon baffles of practical size, or used with horns of reasonable dimensions, begin to show a progressive loss of output; but in good examples this may not fall more than 20 db. at the lower limit of audibility, i.e. some 15 or 30 cycles. Above 5,000 cycles per second response may fall off, but in good examples once again it will be maintained at about the same average level up to perhaps 8,000 cycles, or with the help of a 'tweeter' up to 15,000 cycles. Within this range there will be narrow peaks and troughs in the response curve exceeding 5 db. variation, but these are gradually being reduced to negligible proportions. Thus we see that the modern moving-coil speaker may not depart very far from the ideal of frequency response, and is not now always the weakest link in radio-reproduction.

With this improvement, more attention is being directed to other speaker defects, such as the difficulty of ensuring uniform

sound-distribution within a room, the production of added harmonics by the vibration of the speaker cone, and the true reproduction of transient wave forms. The latter phenomenon is one which arouses controversy, since it is not readily susceptible to measurement. It is realized that frequency and amplitude response do not entirely represent the performance of a speaker, and that when both these are linear a transient wave form may become distorted. Transients, or impulses of too brief a duration to exhibit a definite frequency, play an important part in music, where they convey 'attack' and the peculiarities of certain instruments, such as *pizzicato* strings. It seems that these should be reproduced with unchanged wave form if possible, and that in this respect the ear may possess some degree of phase appreciation. It can be argued that, since any wave front can be analysed into sinusoidal components, it will be truthfully reproduced if these are retained up to a high enough limit, in which case transient and high-frequency response are bound up together, if not actually identical. Other workers prefer to attack this problem from the point of view of the behaviour of speaker cones under shock excitation. Work is taking place to measure this, in the hope that further improvement will eventually become possible.

It has been found that transient response and the ability to distinguish individual sounds from others are improved by high damping of the cone. This can be best achieved by the use of a horn, with the greater air loading it provides. Also, since a horn-loaded cone need vibrate with less amplitude to deliver a given power output, increased magnetic damping can be provided. Electrical damping of the cone is increased when the external circuit resistance is low, just as a meter movement is damped when its terminals are short-circuited. The use of a low-impedance power stage is therefore likely to improve the reproduction of transients. Opinion is therefore developing in favour of the horn when highly realistic reproduction is aimed at, whilst the baffle speaker remains the most convenient when extreme bass is important, or when space will not permit the use of a horn of adequate size.

EXAMINATION QUESTIONS

1. Give a diagram of a low-frequency transformer-coupled amplifier. What tends to limit the amplification of such an amplifier (*a*) at low frequencies and (*b*) at high frequencies?
City and Guilds of London Institute. Preliminary Exam. 1936.

2. Explain how sound is propagated through a medium, and state what characteristics of the wave affect the aural sensations of pitch, timbre, and intensity. *Institute of Wireless Technology. June 1934.*

3. Describe, with sketch, the construction of a moving-coil type loud speaker. Indicate in a diagram the direction of magnetic flux in the magnetic system. On what does the force applied to the cone depend? *C. and G. of L. I. Preliminary Exam. 1935.*

4. Compare the various types of microphone used in broadcast transmission, showing how they differ in frequency response and directional effects. Give the physical phenomena underlying the action of each type mentioned.

5. What is meant by the undistorted power output of a valve? A given pair of valves are used in the output stage of a radio-receiver. Compare the undistorted output to be expected if they are operated under the following circuit conditions:

(*a*) in parallel, with class A biasing;

(*b*) in push-pull;

(*c*) in class B push-pull;

assuming optimum coupling and voltage conditions to suit each case. Compare the distortion heard in each case when the limit of power output is approached.

6. Why cannot the correct grid bias for a power valve be obtained directly from the anode current-grid volts characteristic? What other characteristics are necessary for its estimation?

7. Explain the terms 'load line', 'power triangle'.

8. Illustrate with diagrams as many methods as you can for connecting the loud speaker to the power valve, and explain the reasons for their adoption.

9. If a loud speaker has an impedance of 1,000 ohms at 200 cycles rising to 2,000 ohms at 400 cycles, what transformer ratio would you select to couple it to a power valve of 6,000 ohms alternating-current resistance?

10. The reproduction from a pentode output stage is apt to sound shrill. Why is this, and what steps can be taken to remedy it? Do you know of any circumstances in which this defect might serve some useful purpose?

11. Compare the various methods by which grid bias can be provided for a power-amplifying stage. What steps are taken to maintain the bias constant when necessary?

12. State the three most important characteristics of a horn-type loud speaker. What is their effect upon the performance of the speaker? *I. W. T. June 1937.*

13. A low-frequency transformer has a primary inductance of 20 henries and a turns ratio of 1 : 3. This transformer is connected in the anode circuit of a valve having an amplification constant of 6 and a plate resistance of 10,000 ohms. Neglecting the distributed capacities, calculate the amplification of the complete stage at a frequency of 1,000 cycles. *I. W. T. June 1937.*

14. Describe the method of amplification known as quiescent push-pull, explaining its advantages and disadvantages.
Grad. I. E. E. 1935.

CHAPTER XIII

PROPAGATION OF WAVES THROUGH SPACE

In Chapter I it was stated that two of the most important problems in radio-telegraphy were, firstly, the propagation of wireless waves of various frequencies to great distances over the earth's surface, and secondly, the nature and origin of the atmospheric disturbances which interfere with long-distance communications and, in the tropics, make communication impossible for hours at a time.

Unfortunately, our knowledge of the physical conditions in the upper parts of the atmosphere which are traversed by wireless waves is meagre; and also the accurate measurement of the strength of the electric or magnetic fields due to the waves at a distant receiving station presents difficulty, especially in the case of short waves. Hence it is only possible to state the known facts which a theory of wave propagation must explain, and to advance hypotheses by which they may be explained, bearing in mind that any hypothesis may have to be modified in the face of new data.

In approaching the question of wave propagation, it is necessary to draw a distinction between the propagation of waves over distances short enough for the earth to be considered as a plane, and over longer distances in which the curvature of the earth must be taken into account.

Dealing first with short-distance communication, it is possible to show that a radiating earthed aerial may be regarded as a Hertzian oscillator. On this assumption it can be shown that the radiated field at a point a few wavelengths distant from the aerial is proportional to the product of the effective height of the aerial (i.e. half the length of the equivalent Hertzian oscillator) and the current in it, and is inversely proportional to the distance from the transmitter and to the wavelength emitted.

In 1905 this formula was tested for short-distance transmissions with short waves by Duddell and Taylor, who found that the inverse distance law held accurately up to about 60 miles if the transmission took place entirely over sea. Over land, however, they found discrepancies at distances as short as 1,800 metres. The intensity of the signals received over land

was less than that indicated by the formula, and it was evident that the ground was exercising an absorbing action on the waves. Austin, in 1911, carried out a large number of further experiments on the strength of received signals. He compared the strength of received signals at 83 km. from the transmitter for waves of 1,000 metres and for 3,750 metres. In the case of the shorter waves the experimental value of the electric force found by him was about one-fifth only of the calculated value, while with the longer waves the experimental and calculated values agreed within the limits of experiment. The attenuation of short waves when passing over water and land was studied in America by Bown and Gillett, who mapped out the topography of a district for different directions around two American broadcasting stations.

The results of these and of more recent experiments appear clearly to indicate that dry ground acts practically as a non-conductor to short waves of high frequency, while it acts as a comparatively good conductor to long waves. Owing to the high resistance of the earth, the shorter waves are bent over towards the earth and the energy contained in them is quickly absorbed. Since the electrical conductivity of sea-water is much greater than that of land, the absorbing action is less in the case of propagation over the sea.

The next important point to notice is the shielding or screening action of mountains, which was first investigated by Admiral Sir Henry Jackson. He found that mountains or masses of rock, whose sizes were comparable with the wavelength of the transmitted signals, almost completely shut off the waves from a ship just under the mountain, while when the ship moved out to sea, the signals returned to their former strength. These observations, which are confirmed by later experiments, clearly show that the mountain casts, as it were, a wireless shadow, and that the waves do not pass through it but are diffracted by it in a manner analogous to optical diffraction. It is to the screening effect of hills or tall buildings that the occurrence of what are known as 'blind spots', of which much has been heard since the introduction of broadcasting, may be attributed. Another cause of such 'blind spots' may also be traced to the presence of stretches of land of especially high resistance, in crossing which the waves may be rapidly absorbed.

Propagation of this type varies uniformly with wavelength,

the absorption from all causes tending to increase as wavelength is reduced. We can explain this jointly, by the increased absorption due to most materials at high frequencies, and by the larger number of wavelengths separating transmitter and receiver, which, assuming a given attenuation per wavelength travelled, will clearly result in weaker signals as this is reduced. The ultra-short waves below 10 metres, now finding important application for television broadcasting and mobile services, show the effects of shielding very markedly indeed. Their propagation at short distances is very 'optical' in nature, being not unlike that to be expected from a powerful searchlight placed at the transmitting station. Solid objects cast sharp shadows, and blind spots are pronounced. The signals are strong within sight of the transmitter, but fall off directly the horizon is passed, and refraction is only able to extend their range some 20 per cent. beyond the optical horizon. Attenuation is then very rapid and the waves may soon become inaudible. They are thus unable to follow the curvature of the earth in the manner of longer waves.

If, now, we turn to long-distance transmission, the question at once arises as to whether diffraction effects are sufficient to explain how the waves can get over the huge mountain caused, as it were, by the curvature of the earth. After long debate among mathematicians it was accepted that pure diffraction cannot explain the long ranges achieved by radio-telegraphy on any wavelength, and therefore we are led of necessity to consider other theories of propagation.

It was early found that the transmission range possible with similar powers was very different for waves of intermediate length (say, 250 metres to 2,000 metres) during the day and during the night. A ship fitted with an ordinary $1\frac{1}{2}$ Kw. transmitting set can, for example, maintain steady communication at a distance of perhaps 150 to 200 miles by day on a wave of 600 metres. After dark, however, communication has been effected by similar ships at over 1,000 miles. Similarly, the Scottish broadcasting stations are heard with the greatest difficulty in London in the daytime, even with the most sensitive receiving apparatus, while after dark their signals may be strong. The signals, however, at night, in the case of such communications as the above, are very variable in strength, and what are known as 'fading' effects occur. Thus, although the

transmitting and receiving apparatus may be working perfectly efficiently, the signals, either comparatively suddenly or gradually, fall away to practically zero intensity and then increase again to more than normal strength.

With longer waves the average intensity of signals over a given range for night and day transmission is much more nearly equal and the 'fading' effects at night much less noticeable. On a 6,000-metre wavelength sunrise and sunset effects were observed daily between the Marconi stations at Clifden (Ireland) and Glace Bay (Canada). For a short period at these times the signals decreased to a strength below that at which satisfactory communication was possible. The mean strength of signals during the night and day was, however, equal, and the same is true for longer waves. The general result of observations taken over a long period is that for long-distance transmission signals are stronger during winter than during the summer, but that the relative amount of increase during the winter may vary from year to year.

If we turn now to short-wave transmissions and, in particular, to the experiments with 100-metre waves carried out by Marconi, the following facts appear. Signals as a rule were strong when transmitted over great distances as long as darkness extended the whole way between the stations, while during the day the signals were weaker and varied inversely in proportion to the mean altitude of the sun above the horizon. Lengthy fluctuations of strength, such as the sunset effect with longer waves, were less noticeable. Finally, it is necessary for any theory of propagation to explain the sudden variations in the apparent bearing of distant stations observed with radio direction-finding apparatus.

The factor which enters into all the phenomena of short-wave propagation appears to be the presence or absence of sunlight. It is inconceivable that sunlight itself can change the transmitting properties of sea or land over which the waves are passing at a rate sufficiently rapid to cause the sudden variations of intensity and direction which have been referred to, and so we must assume that it is the state of the upper atmosphere which is varied by the sun. It would appear also, in view of the absorbing action of the ground for short waves, that in their case practically the whole of the radiation is propagated through the upper atmosphere, while in the case of long waves, the

greater part of the radiation is propagated along the surface of the earth. In the case of intermediate waves part of the radiation is propagated through the upper atmosphere and part horizontally along the earth's surface, and it has become general to refer to these two forms of propagation as the Space Wave and the Surface Wave respectively. If we assume that waves

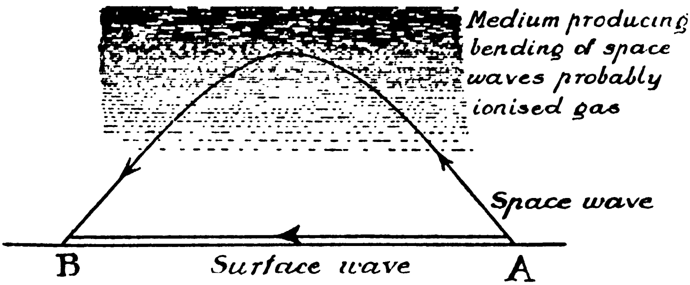


FIG. 130.

which are propagated upwards return again to the earth, then from some cause or other refraction or reflection effects must take place which cause such return, as shown in Fig. 130, where the waves from A are assumed to reach B by the path indicated. An explanation of such bending is found in the ionization of the atmosphere due to the sun. During the daylight hours the ionized portions of the atmosphere extend comparatively near to the surface of the earth: after sunset recombination among the ions takes place in portions of the atmosphere where the gas pressure is fairly large; but there is reason to believe that the upper atmosphere at a height of 70 to 100 kilometres is left in an ionized state. Evidence of the presence of such a conducting shell round the earth can be found in the variation of terrestrial magnetism and in auroras, and is now substantially proved. With waves of intermediate length the energy propagated upwards through the atmosphere as space waves is comparable with that propagated along the earth as surface waves. During the daytime, the space wave is practically wholly absorbed, while at night it may be as strong or stronger than the surface wave, according to the amount of attenuation which the latter experiences.

Many of the 'fading' effects and the sudden variations of intensity met with at night can be explained, qualitatively at

any rate, as interference effects between the space waves and the surface waves. For if the space waves have to travel upwards and then be deviated down again, their path may be different by a few wavelengths from that of the surface waves, and the two groups of waves will be in a position either to add their effects together, or to oppose each other. Much direct experimental evidence of such interference has been obtained by Appleton. Also, it is probable that the path taken by the space waves will vary from moment to moment, owing to changes in the electrical condition of the atmosphere, and that changes in the direction of the layers of ionized gas in the upper atmosphere during the night may contribute to the production of 'fading' effects.

An explanation of the night variations observed with direction-finding apparatus is also found in the combined effects of space and surface waves. The magnetic field due to the surface wave may be supposed to be horizontal, since any vertical component would be absorbed by the earth. This horizontal component may also be assumed to be at right angles to the direction of propagation of the waves, except in so far as it may be affected by irregularities of the ground or changes in the earth's conductivity. The space wave which travels upwards, however, need not have its magnetic field horizontal, since there is no absorption of any vertical component which may be present. In any case, when this wave is bent towards the earth again, the direction of the magnetic field will be changed and the wave may reach the receiver with its magnetic field in other directions than the horizontal. Thus, if we take the projection of the magnetic field of the wave on the horizontal plane, this component may not be at right angles to the direction of propagation of the surface waves. The two component magnetic fields in the horizontal will be combined in a resultant field, which will be perpendicular to the apparent direction of propagation of the surface waves, and an apparent error in the observed bearing will result, as shown in Fig. 131. The vertical magnetic component of the space wave need not be considered, as the lines of force due to this component, being vertical, cannot link with a vertical receiving loop. As the direction of arrival of the space wave will be affected by changes which may take place rapidly in the ionization of the atmosphere, the variation of bearings observed may also be very rapid.

The fact also that night variations of bearing occur infrequently over sea up to distances of 90 miles, while over land they occur at distances over 15 miles, may be due to the fact that the attenuation of the surface waves is, as we have seen, very much less over sea than over land.

The method by which the ionization of the atmosphere pro-

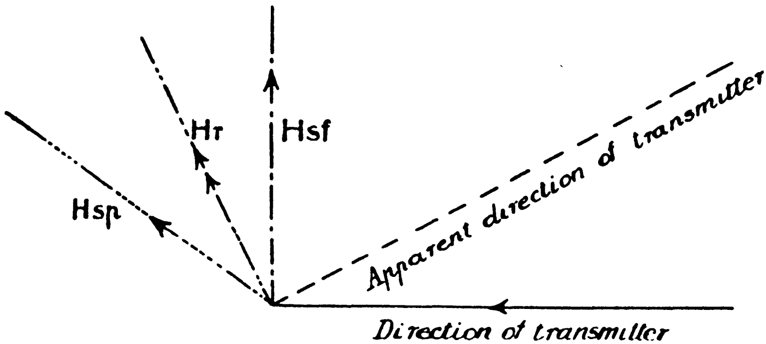


FIG. 131.

Hsp. Horizontal component of magnetic field of space wave.
Hsf. Horizontal component of surface wave.
Hr. Resultant magnetic field linking receiving loop.

duces a bending of the space wave towards the earth must now be considered. When an electromagnetic wave passes through a medium containing free electrons or ions, the ions are set in vibration by the electric force of the wave. As the electrons move they set up small electromagnetic waves out of phase with the original wave. The combination of these two waves has the effect of apparently crowding the original waves together, and making the upper portions of the waves appear to travel with a greater velocity than the lower portion. The result is that the portion of the wave in the ionized gas is bent over towards the earth. The amount of bending produced has been shown to depend on the amount of ionization, and on the square of the wavelength, and to be inversely proportional to the mass of the ions or electrons. The amount of bending therefore depends in no small degree on whether the bending agents are ions or electrons, since the mass of an ion may be some 1,700 times that of an electron. The available evidence points to the bending action being due to electrons.

Another effect must be taken into account, and that is the

loss of energy due to collisions between the electrons and gas molecules present. This loss of energy corresponds to an absorbing action on the waves, and also has the effect of checking the bending of the waves. In order that this effect may be absent, the electrons or ions must be moving in a rarefied part of the atmosphere. Hence, bending can only take place in the upper regions of the atmosphere, where the gas pressure is reduced to such an extent that an electron may move several centimetres before its energy is lost by a collision with a molecule. In this region little attenuation will be suffered by waves travelling through it. Lower down in the atmosphere, however, where the pressure is greater, the presence of ionization may result in considerable absorption of the waves.

It would be expected on this theory that the absorption due to ionization in the atmosphere would be less with short waves of high frequency than with long waves of lower frequency, since with the waves of very high frequency an electron may make several oscillations, even in the denser portions of the atmosphere, before colliding with a molecule. This fact, together with the clearing of the lower atmosphere of free ions after sunset, accounts for the long ranges possible with short waves.

On the theory outlined above it is possible, provided the wavelength is sufficiently short, for the number of electrons in the atmosphere to be insufficient to bend the waves back to the earth again, and thus the waves could escape from the earth's atmosphere altogether. The amount of bending depends on the angle of incidence of the wave with the layer. As the angle at which the waves meet the layer approaches more nearly to grazing incidence, fewer electrons are required to bend the waves to earth again. With waves of, say, 30 metres and below, the effect of the portion of the wave which travels over the ground is very small indeed except very near the transmitter. Up to a distance of about 500 miles from the transmitter the intensity at the receiver is negligible. At a distance of 500 miles, however, the signals become suddenly stronger and detectable up to about 1,500 miles, beyond which reception is uncertain. There is thus a zone of silence round the transmitter of approximately 500 miles radius. This distance is known as the 'skip' distance. The skip distance, of course, varies for waves of different lengths and for different atmospheric conditions. It

may change considerably from hour to hour. The existence of the skip is explained by there being insufficient electrons to bring the wave down again until the angle of incidence becomes that corresponding to the 500 range. At distances closer than this the number of electrons are not sufficiently numerous to bring the waves down, and they are radiated away into space; whilst at longer ranges the waves are effectively reflected back to the ground.

We must now consider the theory in more detail. The definite assumption of an upper ionized layer was made by Heaviside and Kennelly, and hence this has been termed the Kennelly-Heaviside Layer. In recent years the height of this region has been measured by various methods, the principle of which involves the radiation of a wave in the vertical direction from a suitable directional aerial system. A part of this wave will be reflected back towards earth by the ionized layer, and will be received at a receiving station a few miles away. The reflected signal will have travelled farther than one passing over the surface of the earth by a distance approximately twice the height of the reflecting layer, and will therefore arrive at the receiver slightly later than the direct signal. Both incoming signals are recorded by an oscillograph, and from this record the time interval can be measured and the effective height at which reflection takes place deduced.

Measurements of layer height were first made in 1925 by Appleton, employing a somewhat different method. Signal strength was measured at a distance where both direct surface waves and space waves from a transmitter could be received at similar strength. The transmitted wavelength was varied, and it was found that successive minima and maxima of field strength occurred as these two waves became either equal or opposite in phase. This would occur every time the path difference between the surface and the reflected wave equalled an even or odd number of half-wavelengths, and knowing the distance between the stations and the wavelengths corresponding to maxima and minima, the height of the reflecting region can be estimated. Fig. 132 gives an idea of the geometry of either method of measurement. If A and B are the two stations, which are assumed sufficiently near together that the curvature of the earth can be neglected, then the path of the surface wave will be AB. That of the reflected wave will resemble the dotted

curve, but it can be shown that this is equivalent to ACB, and that the height measured will be CD, a slightly greater figure than the commencement of the ionized region. Since the latter is indefinite, however, and has no sharp boundaries, its location can only be found approximately.

In the first method of measurement the distance AB will be very small and the angle ACB acute. We can then say that the path difference between the space and surface waves is practically equal to twice AC, or twice CD, and this is the figure given by the measured time interval between the two signals. An improved form of this method has given excellent information recently in the hands of Appleton and others. Here the transmitted signal consists of brief regular pulses, which leave the receiver unoccupied during the intervals between them. The reflected wave is received during these intervals, and is thus not confused with the other. In the second method of measurement the distance AB is greater, and simple trigonometry must be employed to obtain the height CD, which equals $\frac{1}{2}AB \tan \Delta$.

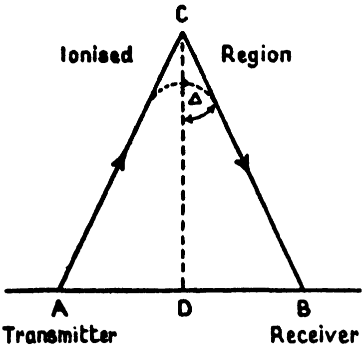


FIG. 132.

As a result of this work an average height for the ionized layer was at first found to be from 80 to 100 km. On occasions, however, anomalous results were obtained, showing reflection at much greater heights. These were interpreted by Appleton as indicating a second ionized region, the average height of which is from 200 to 300 km. Two such regions are now definitely accepted, and are referred to as the lower or E layer, and the upper or F layer, and each of these can be subdivided into sub-regions, E_1, E_2 and F_1, F_2, separated by a less-ionized zone. Very recent work on ultra-short waves shows the presence of at least two more layers, which have been termed C and D layers, and which we shall study later. Signals on wavelengths between about 30 and 300 metres are reflected entirely by the E layer at most times of the year, and for this reason it was the first region to be observed. The F layer reflects shorter waves down to 8 or 10 metres, and was revealed when shorter wavelengths

came to be used in making the experiments, although their behaviour had already shown something of the sort to exist.

The evidence of very extensive propagation measurements and observations on commercial services, coupled with our knowledge of the probable structure of the ionosphere, has lead to a reasonably complete theory by which the rough agreement between ionization and sunlight previously mentioned can be brought nearer to an exact science. It is found, as might be expected, that the lower *E* layer follows the variations in sunlight very closely, and with little time lag, the density of ionization being closely proportional to the altitude of the sun at any given time. The density of the gases composing the atmosphere at this altitude is quite high, and hence recombination of the ions follows quickly upon the removal of the sun's rays. Long waves are unable to pass through this layer and are either completely absorbed by it or largely reflected. Shorter waves may be able to pass through the layer to reach the *F* region, suffering attenuation in so doing, and will be slightly or only partially reflected at favourable states of ionization. Very short waves correspond to frequencies considerably higher than the collision times of the molecules and electrons which compose this layer, and consequently pass easily through it with slight attenuation and negligible reflection or refraction.

Turning to the higher *F* layer, it will be found that at a height of over 200 km. the atmosphere is considerably rarefied, and hence the decay of ionization by recombination of the electrons and molecules after sunlight ceases is quite a slow process. If we realize that several hours may elapse before any large changes in ionization can take place, and that the state of the layer therefore lags considerably behind the progress of sunlight, we find an explanation of many effects which could not be explained when only a single ionized layer was known.

The state of the *E* and *F* layers varies very much with the intensity of sunlight, and hence with the time of day and the season of the year. It has also been found to be affected by sunspot activity, and recent interesting work on this point shows that periods when sunspots are numerous give rise to exceptional levels of ionization. Whilst the graduation from one ionized state to another is clearly gradual and indefinite, it is possible to distinguish certain typical conditions.

The first of these is full daylight, when ionization at all

heights may be expected to be very great. In consequence of this all short wavelengths experience attenuation as they pass through the highly ionized atmosphere, and since it has been shown that this varies as the square of wavelength, it will be best to select very short wavelengths for long-distance communication by day. The limit is set by the fact that the waves must be reflected by the F layer in the state that it may happen to possess at that time of day and year, and so wavelengths below about 14 metres which begin to penetrate the F layer are unsuitable. Broadly speaking, the shortest waves will be used at times when F layer ionization is at its greatest, whilst at periods such as mid-winter, when this is poor, it becomes necessary to employ longer waves. Thus whilst 15-metre waves may be satisfactory in early afternoon in late spring or autumn, it is necessary to go up to perhaps 25 or 30 metres for equivalent results on winter mornings.

In practice the whole of a long-distance radio channel will not often be in full sunlight, and it is necessary to consider the twilight condition also. Twilight is favourable to short-wave propagation, for the ionization of the E layer is not so great as to prevent waves passing through to the F layer, whilst the latter is already highly ionized and therefore favourable to reflection. Attenuation is much less than in full daylight, and so signals on wavelengths between 15 and 45 metres will in general be stronger. When twilight exists along the whole of the great circle between the two stations, attenuation may be very slight, and it is possible for signals to travel several times round the world before becoming inaudible. This gives rise to 'round-the-world echoes' which occur at intervals of about $\frac{1}{7}$ second. They impart a hollow sound to speech, and can cause objectionable distortion to other services. The twilight condition is not therefore always the most satisfactory for communication, although yielding the strongest signals, and a wavelength should be chosen sufficiently long that signals will not pass more than once round the world before becoming absorbed. The shorter waves will be useful for echo observations during this period, and may give rise to astounding ranges, but will probably not be commercially reliable.

At this point we may digress for a moment to deal with the question of echoes in general, namely, signals which reach a receiver at an appreciable time interval after the direct or sur-

face wave has arrived. These fall into three distinct classes, distinguished by the time interval which elapses. The first are quick echoes of short duration, and are due to signals arriving by more than one path through the ionosphere. They are the type mentioned in the measurement of layer height, and occur after very small fractions of a second. They are often multiple and very complex in nature, though fortunately of low intensity.

The second class are those which have just been described as passing completely round the world along a great circle route. They occur at regular intervals of $\frac{1}{7}$ second, several may follow each other at this interval of time, and they are troublesome at times of best propagation.

The third class are termed 'long-delay echoes' or 'Stormer echoes', after the observer who first noticed them. They are rare and weak, but undoubtedly exist, and occur after times as long as 30 seconds from the original signals. They are also more distorted than the two previous types, and more difficult to observe. Beyond a suggestion that these echoes may be due to reflection from some body or strata in outer space, nothing is known of their origin.

A further distinctive propagation condition occurs after sunset and during night-time, but when darkness is not the most complete possible. It will occur in late evening in winter, and during summer nights. Ionization is now slight in the lower atmosphere, and signals very little attenuated. *F* layer ionization will also have fallen off, and very short-wave signals will no longer be refracted back to earth, but will penetrate into space. Hence a lower limit exists at about 20 metres, but waves above this behave effectively. We can distinguish this night condition from a more complete late-night condition occurring in mid-winter. Here ionization is insufficient to return any but the longer wavelengths to earth, and it is necessary to increase wavelength to above about 50 metres before good propagation occurs.

When it is borne in mind that most long-distance communication will take place over paths which are partly in daylight and partly in darkness, it will be seen that the conditions are complex. The best wavelength to employ at any given time must be chosen with regard to all parts of the path, being short enough to give the lowest possible attenuation, but not so short as to escape adequate reflection and refraction at the

F layer. The choice will generally be a compromise, and clearly it will be necessary to vary the wavelength to suit the time of day and of the year. It is not often possible to employ the same wavelength for all the 24 hours between any two distant points, and may not be possible for longer periods than about 4 hours. For this reason several transmitters, or methods for changing wavelength from time to time, are employed on nearly all long-range commercial channels. Very brief listening on any good 'all-wave' broadcast receiver will show that whereas during full daylight or summer the bulk of commercial activity occurs below 30 metres, in winter or at night these wavelengths will be found less active and the bulk of stations will be found on wavelengths from 20 to 60 metres.

It is beyond the scope of this book to go more fully into the study of short-wave propagation under all possible conditions. A method of studying this has been evolved by Eckersley and Tremellen, in which a series of charts are prepared which show the conditions of light or darkness as zones which can be superimposed upon a map of the world. It is then possible to study the conditions through which a signal may have to pass and to select the optimum wavelength more successfully.

We have seen that a typical space wave leaves the transmitter at an angle, persists at this angle until it enters a region in which ionization is increasing, and is then progressively refracted downwards. Three possible cases can now exist: either the wave is not bent sufficiently to return it to earth, or it is bent until parallel to the ionized layer when it proceeds to travel within that layer without further bending, or it is refracted sufficiently to return to earth at a distant point beyond the skip distance. It does this at some descending angle and at a complex polarization which may be mainly circular or elliptical, but will contain a horizontal component.

The first of these cases we can neglect, since no signals reach a distant point, and only ground-wave working is possible. It is illustrated by *a* in Fig. 133. The second is a special case of importance, for once the wave is bent into the plane of the ionized layer it may continue for a considerable distance without very great attenuation, and if it then encounters a region of further increasing ionization such as would occur in passing from night conditions into a dawn area, the wave may be further bent downwards and returned to earth at good strength. This

is one form of distant propagation that not infrequently occurs, and is illustrated by *b* in Fig. 133. The third case mentioned is, however, the most usual one, in which the wave is sufficiently refracted to give the effect of reflection, and returns to earth at an angle similar to that at which it left the transmitter. It is not difficult to see from *c* in Fig. 133 that propagation of this

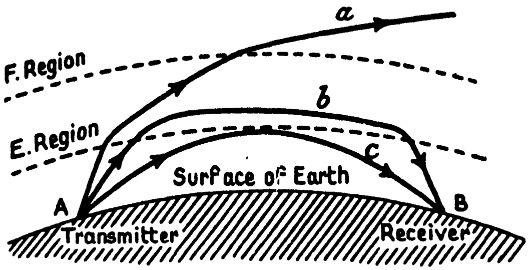


FIG. 133. Propagation of Radio Waves.

kind will be favoured if the transmitter radiates maximum power at an angle to the horizontal, and that the distance at which the waves return to earth for the first time will be greatest if this angle is small. Thus it is found that for the longest ranges in very short-wave working radiation should be concentrated at an angle of from 10 to 20 degrees to the horizontal, and we shall see in the next chapter how aerial systems can be designed to give this effect. Radiation at higher angles is not so effective unless shorter distances are to be covered, as the waves penetrate more deeply into the ionized regions and will be less completely reflected, returning to earth also at a shorter distance from the transmitter. Unless communication by the surface wave is desired, it is waste of energy to radiate parallel to the earth's surface, as such radiation is soon absorbed and contributes nothing to the distance signal. Hence it is avoided in modern aerial systems.

Since any practical transmitter will radiate power at a variety of angles to the horizon, it is clear that waves may strike the ionized regions at a variety of angles, and may reach a distant receiver by a variety of paths similar to the two illustrated. Hence the actual received signal will be the aggregate of many waves arriving by all possible paths. These various components are not necessarily in phase, and their phase relationships will be haphazard and variable. Thus their sum will not always be

constant, and whereas the tendency will be for signal strength to fluctuate about a mean value, a degree of fading is undoubtedly due to this cause.

It is found that when a wave returns to earth at less than a certain critical angle depending upon the conductivity of the soil at that point, it will be largely absorbed; but that when the angle exceeds this the waves are largely reflected and pass back once more into the upper atmosphere. After much discussion and experiment this theory of repeated reflections at the ionized layers and at the surface of the earth is now accepted as the most probable one, and it is thought that waves may reach distant points beyond the range of direct single reflection by a series of these 'hops'. At each reflection the signals lose energy, and so a direct reflection usually to less than 1,000 miles from the transmitter will be stronger than a second- or third-order reflection.

It has also been proved that a degree of scattering takes place when the waves strike an ionized layer, and that this scattering is responsible for weak signals being heard even within the skip area. The combination of scattering and multiple reflections explains the fact that signals are found to be fairly uniform in strength over large areas at a distance from the transmitter, and sometimes at all parts of the earth. Since the received signals are the sum of a large number of waves which have travelled by various paths, their resultant would be expected to be fairly uniform in magnitude. It is at distances just beyond the skip distance where the first reflection only is being received that variations in strength are often most noticeable.

Other forms of fading have been mentioned in earlier paragraphs. In the days before space-wave propagation was understood, the type of fading most noticeable on long- and medium-wave stations at considerable ranges was thought to be due to variations in the attenuation due to the atmosphere and objects on the earth, but the explanation was very unsatisfactory and the problem little understood. Since the discovery of the ionized regions and the measurement of their influence upon propagation, however, fading is very largely explained.

It has just been explained that fading will naturally occur from the combination of various waves reaching a distant point by different paths through the ionosphere. These will combine to produce a resultant which is continually varying in ampli-

tude, phase, and polarization, and in some cases in apparent direction. Taking these factors into account, it has been found possible to reduce fading of this type by combining the signals picked up from vertical and horizontal aerials, having limited directivity. In many cases fading is the result of shifting phase and polarization, and it is found that the horizontally and vertically polarized components of the incoming wave do not fade together, but tend to be complementary. The use of an aerial system which is unaffected by polarization is thus helpful in reducing fading.

A more serious and violent form of fading occurs within an area surrounding medium and medium-short wave stations which is at such a distance that both the surface wave and the space wave can be received. This will be the case just beyond the skip zone of wavelengths which are not too short to enable the surface wave to travel that far, and is most likely to occur near sunset and sunrise, when the skip is shortest and the space wave gaining in strength. In the case of medium-wave broadcasting stations this zone commences at from 50 to 100 miles from the transmitter, and is the form of fading most noticeable in ordinary foreign station listening. This form of fading is due to the combination of the surface and space waves, and since these have travelled unequal distances they will in general differ in phase. There will, broadly speaking, be a single space wave-front having definite phase and polarization, since we are considering a locality too near to the transmitter for any but first-order reflected waves to occur, and these only when near the optimum angle. Should the two waves be nearly the same strength, then clearly it is possible for the resultant field to vary from zero to double the strength of either, according to whether they are in opposite or similar phase. Fading can therefore be very deep, and tends to occur in a moderately slow and regular manner as the path of the space wave slowly changes under the influence of slowly increasing or decreasing ionization.

It is found that signals from any radio-transmitter fade quite differently at points a few wavelengths apart, and that the field at two such points will not often be zero or a maximum at the same instant. Hence if several aerials are erected each several wavelengths away from its neighbour and the resultant signals combined into a common circuit, there is a tendency

for irregularities to cancel out and for the signal strength to be more uniform than at any single aerial. In particular, the resultant signal will very seldom approach zero, as at least one aerial will usually be receiving a substantial signal at any moment. This arrangement is due to Beveridge, and is known as Diversity Reception. It is much used for such work as distant programme relaying and for long-distance telephony channels. The signals are usually combined at audio frequency, separate receivers being used at each point.

The effects of fading are further minimized through the use of automatic volume control in most modern receivers. The effect of this is to automatically adjust the sensitivity of receivers to compensate for changes in incoming signal strength. This device is very helpful, but it should be realized that it cannot overcome cases when the signal fades to practically zero, and also that it does not improve the signal-to-background noise ratio. Automatic volume control can be very usefully employed to assist diversity reception. The control circuits of all the receivers are linked together in such a way that the receiver which is delivering the strongest signal at any given moment will automatically reduce the sensitivity and hence the output from the other receivers to a low level. This prevents those receivers which are not delivering a useful signal at that moment from contributing noise, which would otherwise raise the total noise level.

It is interesting to note that Robinson has also advocated the use of diversity methods in transmission, showing that the field due to a group of well-separated low-power transmitters all delivering the same programme on the same wavelength will tend to be uniform, free from fading, and altogether more consistent than that from a single station of the same total power. It might be thought that the field from the several transmitters, being of haphazard phase, would tend to a zero resultant; but calculation shows that this is not the case. It can be shown that for N transmitters each of power P working on identical wavelengths, the resultant field strength at a distant point is proportional to $\sqrt{(PN)}$, and becomes increasingly constant as N is raised.

Short-wave propagation has now been reviewed, and it remains to consider the effect of modern theories upon longer wave working. Proceeding upwards from the short-wave region, we come to the medium-wave band of from perhaps 100 to

1,000 metres. The types of fading peculiar to this band have just been considered, and at night it can be said that propagation resembles the longer short waves very closely. The space wave may be expected to predominate, but since attenuation of these longer waves in the ionized regions is always appreciable, signals become weaker at great distances. The night range of a medium-wave station is very approximately 1,000 to 2,000 miles, and only under the most favourable conditions of ionization will this be exceeded. It will frequently be less. On certain favourable winter nights in this country, for example, American medium-wave broadcast stations can be well heard, but this is not possible at average locations on the majority of nights.

By day these longer waves resemble the very long waves, and can be considered together with them. Propagation is by a form of surface wave. Attenuation is always appreciable, but decreases with wavelength in the exactly reverse manner to the short waves. We mentioned at the beginning that early workers were puzzled because long waves seemed to follow the curvature of the earth, and this was one of the phenomena that suggested the ionized layer. It is now known that the waves are in fact kept from dissipation into outer space by the *E* layer, which is quite opaque to long waves, and that they propagate by a method which is best expressed as similar to a 'spherical transmission line'. It is beyond the scope of this book to consider the full meaning of this statement, but the long waves should be considered as travelling in the relatively narrow space formed between two highly conductive spheres, namely, the earth and the *E* layer respectively. These prevent the waves spreading in the radial direction, and constrain them to travel freely along the circumference of the inner sphere. Losses are mainly those due to imperfect conductivity and reflection by the earth's surface and by the *E* layer, and would be zero if these surfaces were perfect conductors. The less long waves will show day and night changes, produced by the varying conductivity and reflection of the *E* layer as ionization rises and falls. These effects become less noticeable as wavelength is increased, being nearly negligible at the longest in commercial use. Fading also decreases, and the rapid variety is seldom experienced with waves exceeding about 1,000 metres in length.

The vagaries of the *E* and *F* layer explained most long-, medium-, and medium-short-wave phenomena adequately enough, but

became increasingly inadequate when applied to the short- and ultra-short-wave regions. The persistent work of Watson Watt and the Radio Research Board brought to light some further facts during the period from about 1930 to 1935. One of these was the realization of the subdivided nature of the *E* and *F* layers already mentioned, their division into F_1, F_2 and E_1, E_2, which added detail and accuracy to our theories of reflection and refraction. It was proved that periods of exceptional ionization can occur in the *E* region, termed 'Intense *E*' conditions, which occasionally reflect back to earth waves so short that they would normally penetrate both *E* and *F* regions. It was found that ultra-short waves below 10 metres seem to be returned to earth at ranges such as 1,000 miles when these conditions exist, Watson Watt suggesting that the recent reception of the London television signals in Berlin and vice versa arises from this cause. Increased information was gradually collected concerning the eleven-year sunspot cycle and the corresponding cycle of average ionization density which accompanies it. Years of maximum sunspot activity have been shown to correspond to periods of abnormal *F* layer ionization when signals below 10 metres may be returned to earth at the greatest terrestrial distances. Such a period is expected in 1939, and already some time before that year exceptional ranges are being noted. Amateur 10-metre signals were yielding intercontinental contacts of exceptional strength even during the winter months of 1937, a time of year when this was not to be expected. The frequent reception of American police radio signals between ·7 and 9·5 metres in this country is another example of the exceptional ionization which is preceding the 1939 sunspot maximum, one of unusual intensity in itself. It is unfortunate that these results will probably vanish entirely towards the years of sunspot minimum, and have therefore only a limited value for commercial communication.

By 1935 the existence of a *D* region lying between 40 and 60 miles above the earth was fairly well established. The electron collision frequency is high in this region, and it therefore causes appreciable attenuation of the shorter wavelengths which pass through it. Later work has indicated the region to be fairly uniform and diffuse over its thickness of some 20 miles. Signals can be detected by reflection from all parts of this thickness, but not usually at sufficient intensity to influence communica-

tion services. There seemed no reason at that date to expect the return to earth of ultra-short waves from lower levels than this, although isolated examples of fading and propagation over distances well beyond the horizon demanded an explanation.

Between 1934 and 1937, however, Watson Watt and his assistants have added a further important chapter to this subject, by the identification of some six new layers at lower levels.[1] These lie at approximately 4·7, 5·3, 5·8, 6·4, 6·8, and 8·5 miles above the earth's surface, are comparatively sharply defined, and are surprisingly high in ionic density. Owing to the close spacing of these layers, and the fact that reflection is so perfect that echoes up to the tenth order can be observed, indicating a reflection coefficient as high as 70 per cent., it was not possible to separate them individually in a receiver. A wavelength near 50 metres was reflected back vertically so as to form a reflection pattern, composed of the overlapping in time of these several reflections from the several layers, and a very long period of patient work was necessary before these patterns could be interpreted. In addition to these fairly well-defined low layers, an upper region lying between 20 and 30 miles was also detected, but has not yet been analysed with the same detail into the individual layers which may compose it.

We thus see that the atmosphere is quickly being shown to be highly stratified from the ionic point of view, there being anything upwards of ten regions of maximum intensity lying between 240 miles or more for the *F* layer, down to a mere 5 miles or less. The upper layers which we have studied in outline are responsible for the bulk of long- and medium-wave phenomena which vitally influence long-distance communication. The newly discovered low layers have a most pronounced effect upon moderately short-range communication, and at the shortest wavelengths. They can return ultra-short waves to earth at distances of a few hundred miles, and are probably responsible for the unexpected deviations of these waves from the simple 'optical' surface-wave theory. The fading and freak ranges of signals between 5 and 10 metres and the multiple image effect occasionally observed in television reception seem to be accounted for by their existence.

[1] 'Wireless and the Atmosphere', by Watson Watt, in *Wireless World* of March 5th, 1937.

It is not easy to see from what source these lower layers receive their ionizing energy. One source has been detected in thunderstorms, which are believed to produce quantities of free electrons in the region above them, namely, at the order of 6 miles in height. It has been observed that low-layer ionization is considerably increased by the presence of a storm within 15 miles of the observing station. A thunderstorm may thus affect ultra-short-wave conditions, whilst on the longer wavelengths it gives rise to atmospherics, which, next to irregular propagation are the chief barrier to reliable communication.

We must now turn to this other equally important problem of radio-telegraphy. From 1906 to 1918 over a hundred patents were taken out for devices intended to overcome atmospherics, and hundreds more have been taken out since that date. Nevertheless, it is a matter of agreement among radio engineers that the advance made towards any real mitigation of the nuisance by any devices, other than by directional aerials and sharply selective circuits, is small.

In this country the Radio Research Board decided, when they took up the study of atmospherics in 1920, to leave the problem of elimination severely alone, and to concentrate on the investigation of the nature and origin of atmospherics. For this purpose a specially equipped station was erected under the superintendence of R. A. Watson Watt, and results of the greatest importance have been obtained by him, which will be briefly reviewed. It may be observed, however, that the lead given by this country in concentrating on the fundamental side of the problem has been followed by workers in several other countries.

Until about ten years ago the only measurement on atmospherics which could be made with any possibility of clear interpretation was that of the direction from which the atmospherics appeared to come, and it was to the study of the variations in this predominant direction of arrival that Watson Watt first directed his attention. Before atmospheric investigations were undertaken by the Radio Research Board, the Meteorological Office had arranged for directional observations on atmospherics to be made by the Naval coast direction-finding stations in this country. Between 1916 and 1920 about 15,000 observations of this description were made. The first important fact arising from these observations was that simul-

taneous readings of the direction of arrival of atmospherics at several stations frequently gave intersections at points at which it was known from meteorological data that a thunderstorm was in progress. In this way thunderstorms were located by wireless direction-finding as far away as the coast of Sicily, the Bay of Naples, and Helmond in Holland. Recently, with improved apparatus, a meteorological depression associated with thunderstorms has been tracked, by means of the direction of arrival of atmospherics apparently originating in it, for 2,000 miles from the Hebrides across Europe to the Black Sea.

The data available over a period of years from these short-wave direction-finding stations were examined. When simultaneous observations of the direction of arrival of atmospherics by three or more stations indicated a definite place of origin of the atmospherics, it was found that 94 per cent. of these places of origin were associated with areas in which rain had fallen in the twenty-four hours containing the time of observation. In 15 per cent. only, however, could definite agreement with reported thunder be obtained, although in 10 per cent. more cases the meteorological reports available indicated phenomena usually associated with thunder. More recent observations have shown conclusively that atmospherics are produced by electrical disturbances in the atmosphere, which are in the nature of thunderstorms, but not necessarily accompanied by actual thunder. Atmospheric centres have been located over all the larger land areas where drastic temperature changes and thundery conditions prevail, notably tropical districts in India, Africa, Central Europe, and the tropical zones of North and South America. A seasonal variation in the position of atmospheric-producing zones has been noticed, which is mainly a movement a few degrees to the north during the European summer, and south during the winter. There is also a tendency to maximum activity shortly after noon, local time.

It is a matter of general experience that the effects of atmospheric disturbances are greater with longer than with shorter waves. For that reason, when the study of the direction of arrival of atmospherics was undertaken at the Radio Research Board's Station, the directional apparatus was first designed for 20,000-metre waves. The average direction of maximum disturbance from atmospherics in Great Britain, as determined by observations during 1921, is somewhere between south-east and

south-south-east. Analysis of more recent data regarding direction of arrival of atmospherics shows that the apparent direction follows the sun but with a lag in time, atmospherics arriving from the east about 9 a.m., from the south towards 6 p.m., and from westerly directions in the dark hours. About 8 to 10 a.m. the stream from the west dies out and is replaced by that from the east.

When we come to the study of the nature of atmospherics apart from their origin, it is clear that the difficulties to be faced are great. Let it be supposed for a moment that an atmospheric consists of a complete single wave the length of which is 300,000 metres. Such a wave falling on an ordinary aerial may introduce an electrical impulse lasting only $\frac{1}{1,000}$ of a second. But if the aerial is only feebly damped, the current in it will not necessarily cease at the end of this period. Thus it might be difficult to disentangle the form of the 'free' vibration (i.e. the vibration depending on the constants of the aerial circuit) from the 'forced' vibration (i.e. the vibration depending on the form of the external applied force, in this case the atmospheric). In order to overcome this difficulty it is necessary to choose the constants of the aerial circuit in such a way as to reduce the free disturbance to a negligible amount compared with the forced vibration. This is done by introducing a large ohmic resistance into the aerial circuit, which changes the 'free' vibration into a 'free aperiodic' disturbance lasting only perhaps $\frac{1}{100,000}$ second. Hence for atmospherics lasting longer than this period it can safely be assumed that the electromotive forces in such an aerial circuit accurately follow the form and intensity of the atmospheric.

What is next required is apparatus for recording the values of the electromotive force in the aerial with time, bearing in mind that the phenomenon to be recorded is a transient one lasting only, say, $\frac{1}{1,000}$ second in all. Watson Watt and E. V. Appleton independently suggested that a suitable instrument for this purpose would be a sensitive cathode ray oscillograph working at a low voltage. We have already seen when studying valves what a cathode ray tube comprises, and that its rapid distortionless response renders it eminently suitable to work of this kind. The afterglow of the fluorescent material enables the observer to see any pattern traced by the spot of light on the screen, even though its movements may actually be extremely

rapid. Before reaching the screen the beam of electrons passes between two pairs of plates at right angles to each other, shown as *mm* and *nn* in Fig. 134. If the electromotive force due to the electric field of the atmospheric is applied to the plates *mm*, the spot will be drawn out into a straight line. In order to obtain a graph showing the variations of this E.M.F. with time, a

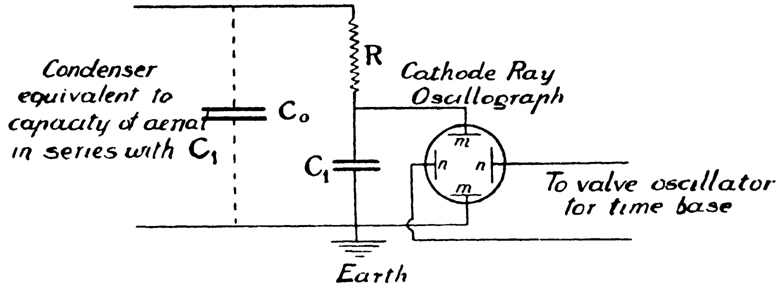


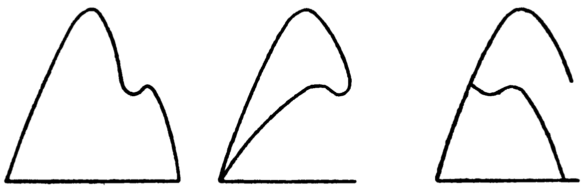
FIG. 134.

voltage is applied at right angles to that of the atmospheric by connecting *nn* to a low-frequency valve oscillator.

This oscillator was at first made to produce a pure sine wave over a wide range of frequencies. Before the application of the atmospheric to the plates *mm*, the beam of electrons will be drawn out into a horizontal line, whose length can be varied by adjusting the amplitude and frequency of the oscillating valve. The spot of light on the screen is really executing a rapid motion along a horizontal line. The spot starts from rest, increases in speed until the middle point is reached, then slows down until it reaches the right-hand end of the line. It then performs a similar motion from right to left, and so on. If the voltage from the atmospheric is applied to the plates *mm*, producing a vertical deflexion, we shall have the base line continually retraced while the atmospheric persists, and, as a result of the combination of the two deflexions, we shall have a curve traced out in which the abscissae measure time, while the ordinates are proportional to the instantaneous voltage produced by the atmospheric.

The form of the curve is complicated by the fact that for part of the oscillation the time is measured from the right and for other portions of the oscillation it is measured from the left. This difficulty has been removed by the design of a linear time base in which the spot is made to oscillate comparatively slowly and

with uniform speed from left to right, and to return very rapidly indeed from right to left, so that the potentials applied to the spot are in the same direction. This is the 'saw-tooth wave form' used in television scanning. The shape of the same electrical discharge form is shown in Fig. 135 as recorded (*a*) by a linear time base, (*b*) by a sinusoidal time base. The periodicity



(*a*) (*b*) (*c*)

FIG. 135.

of the linear time base is assumed to be the same as the duration of the discharge: if it were less, the form might be similar to that shown in Fig. 135 (*c*).

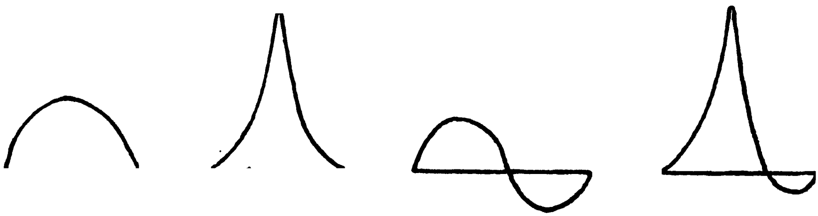
Fig. 134 indicates how the oscillograph is connected to the aerial circuit. It is, of course, impossible to connect the oscillograph across the whole of the aerial: instead, it is connected across a condenser C_1, introduced into the aerial circuit, which is assumed to have a capacity C_0. These two capacities being in series, the voltage applied to the oscillograph plates due to an electrical potential of E volts per metre acting on an aerial, whose effective height is h metres, is given by

$$V = \frac{C_0}{C_0+C_1} hE.$$

The deflexion obtainable with a Western Electric tube is about 1 mm. per volt. As the condenser C_1 must not be too small, it is necessary to introduce in practice one or more stages of amplification between the aerial and the oscillograph. The deflexion per condenser volt is then 1·5 cm. instead of 1 mm. In the work carried out by Watson Watt many thousands of atmospheric forms have been observed, both in this country and in the tropics, the peak voltages being read from a scale on the fluorescent screen itself.

The general shapes of the most commonly occurring forms are as shown in Fig. 136, (*a*) and (*b*) representing what may be termed aperiodic disturbances and (*c*) and (*d*) quasi-periodic

disturbances. The disturbances of rounded type (*a*) and (*c*) were found to be 2·43 times as frequent as the peaked type (*b*) and (*d*). The most frequently recurring duration of the aperiodic class was found to be 1,250 micro-seconds (·00125 second) as measured by the amount of the time base they occupied, but values as low as 100 micro-seconds and as high



(*a*) (*b*) (*c*) (*d*)

FIG. 136.

as 55,000 micro-seconds ($\frac{1}{20}$ second) have been observed. The mean field strength is about 0·125 volt per metre. In many cases the rates of growth and decay of the atmospherics were similar, but in a few cases the rate of growth was too rapid to be observed. In much previous theoretical work it has been frequently assumed that the period of growth was infinitesimally short. This assumption would appear to be entirely incorrect in 80 per cent. of the cases observed.

In the case of quasi-periodic disturbances the most frequent duration was about $\frac{1}{500}$ second. The half-wave of greater amplitude occupied about the same time as the second half-wave in only a small number of cases observed, and on the whole it was found to last 1·7 times longer than the second half-wave. Actually the most common type observed in Great Britain was that shown in Fig. 136 (*a*). One of the outstanding features of the results is the immense strength of the disturbances. The average field of an atmospheric is found to be $\frac{1}{8}$ of a volt per metre, while the field strength of a strong trans-atlantic signal is perhaps only 50-millionths of a volt per metre.

In many cases it was found that the wave forms observed had high-frequency ripples impressed upon them. The periods of these ripples are very near the longer wavelengths used in radio-telegraphy and their amplitude was many times stronger than the strongest signals. An atmospheric wave form of this type is shown in Fig. 137. The forms showing these high-

frequency ripples are far more common in the tropics than elsewhere. Also, in the tropics the number of discharges occurring per second is far greater than in northern latitudes. It is therefore natural that long-wave stations should experience most atmospheric interference in these districts. On shorter waves, however, the case is different, as stations situated in the tropics may be fortunate in that the major local atmospheric centres may lie within the skip distance of the wavelength used and will not be heard. As very short waves are approached therefore stations in temperate climates may be the worst affected.

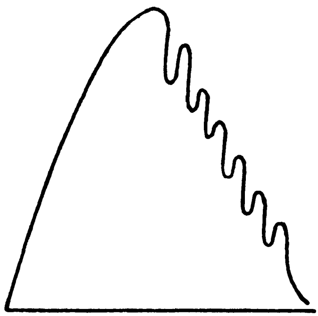


FIG. 137.

Since atmospherics are of fairly long duration, their fundamental frequency will lie within the long-wave region, where they are most troublesome. But we know that a single pulse of energy, particularly if it has a steeply rising wave front, can be analysed by the Fourier series into components covering a very wide frequency spectrum. This will include the highest radio-frequencies at proportionately decreasing intensity, but so enormous is the initial energy of atmospherics that appreciable disturbances are produced down to 12 or 14 metres. They are not serious as a rule below 30 metres, the worst short-wave region being that from 100 metres upwards, often useless for long periods in the tropics except at local range. It is this fact that mainly prevents the full development of a broadcasting service in those countries.

The almost infinite frequency spectrum of many atmospherics explains, of course, the extreme difficulty of their complete elimination in the receiver. The most successful general methods lie in the use of the maximum possible selectivity, thus keeping out interfering energy on all but the wave band in actual use, and by the help of directional aerial systems which are insensitive in the direction of worst atmospheric interference. More elaborate methods have generally made use of some form of balancing, whereby the atmospherics are separately amplified and reversed in phase. Then they are combined with the incoming signal to neutralize their own counterparts. Such systems are

ingenious and superficially possible, but are so difficult to carry out in practice that none have met with complete success. A recent line of attack, termed in America the Lamb noise filter, employs a drastic and quick-acting form of automatic volume control. This is operated by atmospherics and noise above a certain level, which must exceed the signal level, and thus any pulse which is decidedly stronger than the signal will very quickly and briefly 'kill' the receiver. This must occur before the circuits which carry the signal have time to function, a slight time lag being introduced for that purpose. Then the atmospheric automatically causes a brief gap in the reception, during which time its own noise is suppressed. These gaps are too brief to be heard by the ear, but they prevent the atmospheric from being heard and from shock-exciting the loud-speaker into violent vibration.

On the ultra-short waves atmospherics are negligible, but their place is often taken by man-made electrical noise, much of which is caused by motor ignition systems. This is often most troublesome near towns, and can be minimized by the same methods which have just been outlined for atmospherics.

Man-made interference is on the increase with the growing use of electrical appliances of all kinds. Since complete elimination at the receiver is so difficult, it is much better to suppress this interference by suitable filters at the source. Legislation is being introduced by most countries to prevent the sale of equipment unless fitted with adequate silencing devices, such as condensers across all sparking contacts and radio-frequency chokes in mains supply leads. With the development of ultra-short-wave services it may be necessary to fit all cars with devices to prevent radiation from the ignition system. A damping resistance of about 15,000 Ω in series with each sparking-plug, together with condensers across the contact breaker and dynamo brushes, form an adequate cure in most cases.

Interference reaches many receivers over the supply mains by which they are fed. An improvement is possible in this case by inserting a low-pass filter in the supply cable near to the set. Interference may be injected into the mains by offending equipment, when it may be radiated again from the mains at distant points. It may therefore be helpful to instal a receiving aerial well clear of buildings and electrical wiring or apparatus, and to

join it to the set by a fieldless transmission line. 'Anti-interference aerials' of this kind are now much used for domestic broadcast reception, and the principles by which they work will be described in the next chapter. It is generally recognized, however, that the best point at which to eliminate interference due to electrical equipment is at the source.

EXAMINATION QUESTIONS

1. Short waves transmitted over a long distance in a north-south direction behave differently from waves transmitted over a similar distance in an east-west direction. Explain the reason for this behaviour. *Institute of Wireless Technology. May 1936.*

2. Write a report of about 150 words on special features of the propagation of ultra-short waves and their application to television broadcasting. *I. W. T. May 1935.*

3. What is the chief source and the nature of atmospherics ? Why cannot they be tuned out or completely eliminated ?
I. W. T. November 1934.

4. What do you know of wireless echoes ? Describe the three types that can be observed, and explain how they are thought to originate.

5. Describe the structure of the ionized regions of the upper atmosphere, and outline the effects which they exert upon wireless waves of various frequency.

6. Write a short account of one of the following:

(*a*) A magneto striction oscillator.
(*b*) The Kennelly-Heaviside layer.
(*c*) A super-regenerative receiver.

Note: (*a*) is not dealt with in this book. For (*c*) see Chapter X.
A. M. I. W. T. June 1937.

7. Draw a diagram giving the location of the various ionized layers in the upper atmosphere. Explain how the virtual heights of these layers may be measured. *A. I. W. T. June 1937.*

8. How can the reflection of radio waves from the ionized regions of the upper atmosphere be observed and measured ? What type of equipment is suitable for such work ?

9. Describe equipment capable of measuring the wave form of atmospherics. What is the nature of the results obtained ?

10. What do you understand by (*a*) the *E* layer, (*b*) the *F* layer, (*c*) skip distance?

11. Explain how refraction of radio waves can arise in the upper atmosphere, and by what mechanism this can result in the bending or total reflection of the waves.

12. Explain how it is that long radio waves can travel over the greatest terrestrial distances. Why do they not pass into space and become lost?

CHAPTER XIV

DIRECTIONAL RECEPTION AND THE AERIAL SYSTEM

THE aerials we have hitherto considered have been assumed to radiate equally well in all directions and to be capable of receiving equally from all directions, except in so far as the screening effect of buildings, trees, hills, &c., prevents the waves from reaching the aerial. If we consider a vertical receiving aerial such as was used by Marconi in his earliest experiments, and if

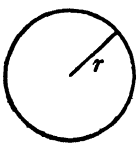


FIG. 138 (*a*).

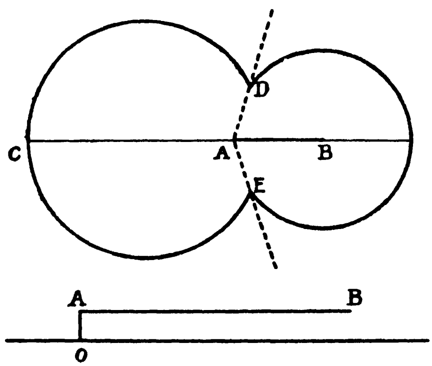


FIG. 138 (*b*).

the intensity of the signals received from a station be represented by a line of length r, then the signal intensities in the receiving aerial for the different equidistant positions of the transmitter will be represented by the radii of a circle of radius r as shown in Fig. 138 (*a*). In other words, it can be said that the polar curve for a vertical aerial in the horizontal plane is a circle.

Aerials can be designed, however, whose polar curves are not circles. Such aerials have directional effects and receive signals more strongly from one direction than from other directions. The earliest type of directional aerial had a long flat portion, the length of which was of the order of ten times the height of the vertical portion. This type of aerial gives a polar curve such as is shown in Fig. 138 (*b*), where OAB represents the configuration of the aerial. It will be seen that, of waves from equidistant transmitters, those arriving from the direction CA

develop the greatest E.M.F. in the aerial, while those arising from the direction *AD* or *AE* give the least signal intensity. The intensities of these signals are represented by the lines *AC*, *AD*, *AE*.

In this chapter we shall deal firstly with the application of directional aerials in the reception of signals. As stated in

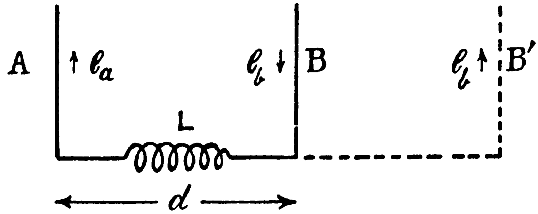


FIG. 139.

Chapter I, the object of such aerials is either to provide a means by which the direction of a transmitting station can be determined, to improve signal strength, or to eliminate interference. The directional effect of the long Marconi aerial was on the whole not very great, and the intensity of the signals from the minimum direction was considerable. One type of directional aerial used to-day is based on the directive properties associated with reception by closed loops or coils of wire known as 'frame' aerials. The directive properties of such loops depend on some of the fundamental facts of wave motion.

Suppose we have two exactly similar vertical aerials A and B, separated by a distance d, and joined at their lower ends, as shown in Fig. 139. Suppose also that an electromagnetic wave excites oscillating voltages in the aerial A of value e_a at a particular instant, and that an oscillating voltage of value e_b is excited by the wave at the same instant in the aerial B. The potential difference across the coil L will be $e_a - e_b$. From the consideration of wave motion given in earlier chapters, it is clear that the potential excited will depend on the distance between A and B and on the direction of propagation of the waves. If the direction of propagation of the waves is *parallel* to the plane containing A and B, and if A and B are exactly a wavelength apart as shown in Fig. 140 (i.e. $d = \lambda$ for the particular incoming wave considered), the potentials excited in A and B at any instant will be equal in magnitude and direction, and therefore no difference of potential will be created across L.

When the distance between A and B is reduced below a wavelength, a difference of potential will be set up across L which will be a maximum when $d = \lambda/2$ and will decrease as A and B are brought nearer together. With suitable amplifiers the oscillations can be detected, however, even if A and B are only a few feet apart. On the other hand, when the incoming wave is

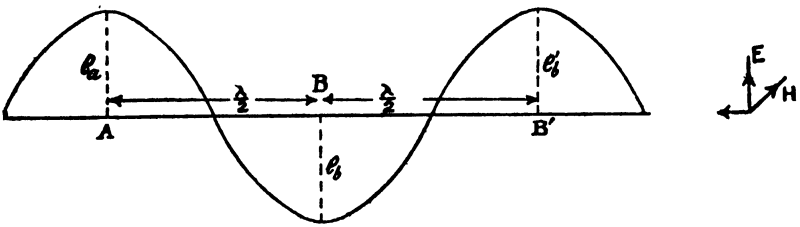


FIG. 140.

propagated in a direction at *right angles* to the plane containing A and B, the potentials due to the oscillations at any moment in A and B will be equal in magnitude and direction, whatever the distance between A and B, and therefore the potential across L will always be zero.

For other directions of arrival the potential will be between zero and the maximum value found when the direction of propagation is parallel to the plane containing A and B.

Thus the polar curve for such an arrangement of aerials is that given by Fig. 141, where r represents the intensity of signals due to waves arriving in a direction making an angle α with YOY', the line at right angles to the plane of the aerials.

It has been found in practice that the effect of such a pair of aerials is increased by joining their upper ends, so that reception takes place on a closed rectangle as shown in Fig. 142. We have seen that the intensity of the received signals depends on the length of the lower side and also on the height of the vertical sides. Before the invention of valve amplifiers, only loops of very considerable size could be employed, but with powerful multivalve amplifiers loops of only two or three feet give good results over long ranges.

We have deduced the action of the loop from consideration of the effect of the electric force in the wave, which we assumed to be vertical. We have seen in Chapter XIII that this assump-

tion is in most cases not strictly true, and the advancing wave front is inclined to the vertical. Whatever the direction of the electric field, the magnetic field due to the wave must be perpendicular to the electric field, and the direction of propagation of the wave must be perpendicular to both. The effect of the loop can be obtained equally well from considerations of the

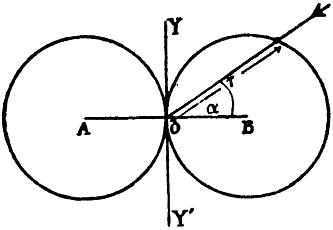


FIG. 141.

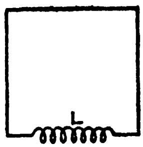


FIG. 142.

potentials produced by the linkage with the loop of the lines of force due to the magnetic field of the wave.

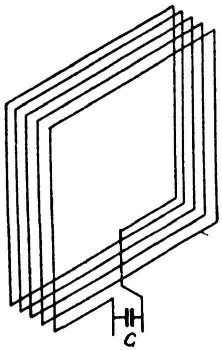
If the plane of the loop is parallel to the magnetic field there will be no linkage, and the received signal is zero, while a maximum signal will be obtained when the magnetic field is perpendicular to the plane of the loop and the greatest linkage of lines of magnetic force takes place.

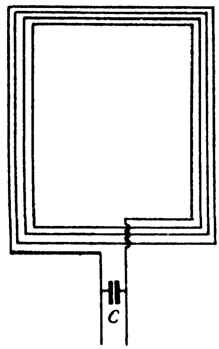
Thus, if signals are received in a vertical loop which can be rotated until the sound in the headphones due to the signals vanishes, then, in general, the direction of the transmitting station will be in the direction of a line drawn at right angles to the plane of the loop. There will, however, be an ambiguity of 180° regarding the side of the loop on which the transmitting station actually lies. Thus, if the signals from a station are zero when the loop is placed at AB, as shown in Fig. 141, the transmitting station may be situated either along OY or OY'.

There are four practical systems of direction-finding in use to-day, they are known as the single-coil, the Bellini-Tosi, the crossed-coil or Robinson system, and the Adcock system. Each of these will now be considered shortly. It can be shown, however, that each of these reduces in principle to the single loop which we have just considered.

The Single-coil System. In order to obtain a suitably strong signal in a loop which is of convenient size for rotating by hand, it is necessary to use several turns of wire instead of

a single turn. Also, in order to tune the loop to the received waves, a condenser is introduced into the lower limb. The oscillating potential developed across the condenser is applied to the grid of the first valve of an amplifier. It can be shown that these potentials vary with the orientation of the loop in the same manner as the electromotive force induced in it by





(a) FIG. 143. (b)

the incoming waves. The turns used in making up the loop are either spaced in equally dimensioned loops in nearly parallel planes (box-type coil), as in Fig. 143 (*a*), or are wound spirally in the same plane (pancake coil), Fig. 143 (*b*).

The potential difference developed across the condenser C is proportional, in the first case, to the number of turns making up the loop multiplied by the area of a turn, and in the second case to the number of turns multiplied by the mean area of the loop.

To obtain the direction of a given transmitting station a horizontal circular scale divided into degrees is attached to the moving loop, and so adjusted that a fixed pointer indicates zero or 180° when the plane of the loop is east and west (i.e. the perpendicular to its plane points north and south). The loop is now rotated until the signal from the transmitting station disappears, and the reading of the pointer gives the bearing of the station from the north. By rotating the frame through 180° a second reading is obtained and the mean of these gives the observed bearing, but with an ambiguity of 180°, as already explained. To find the position of the transmitting station a second directional receiving station situated at some distance

from the first may be used, and the bearings obtained from the two stations are plotted on a suitable chart. The point of intersection of the two lines gives the required position.

There are two instrumental errors which are more or less common to all the systems, but which are most easily understood when considered in connexion with the single loop. The first of these errors is due to the fact that the leads of the whole apparatus may pick up a certain amount of energy from the incoming waves independently of the orientation of the loop. While this energy may be quite imperceptible, compared to the intensity of the signal, in the maximum position of the loop, it may be sufficient to cause an audible signal when the loop is in its minimum position; hence, instead of a sharp zero being obtained, a more or less blurred minimum will result. To avoid this, the receiving apparatus is encased in an earthed metal box to screen it from outside oscillations, and the leads connecting the apparatus to the loop are similarly encased in metal sheathing.

Secondly, one side of the condenser C of the loop is connected to the grid of the first valve, while the other side is connected to the filament, with which are associated the somewhat bulky filament batteries and high-tension batteries. High-frequency currents will tend to flow to earth from the two vertical sides of the loop through any capacity present, and since the apparatus connected on the filament side of the valve will have a greater capacity to earth than the grid of the valve, these capacity currents will be unequal. The effect of this inequality will be to develop a potential difference across C which may produce an ill-defined or incorrect zero. This error is known as *antenna effect*, since the currents producing it are due to the vertical parts of the loop acting as ordinary aerials. The currents produced in this way are non-directional and are superimposed on the currents picked up by the proper directional action of the loop. The result is that the minima obtained on rotating the loop are found not to be exactly 180° apart. A satisfactory method of overcoming this defect, in the case of a single coil, is to introduce a compensating condenser between the grid of the first valve and earth, so that the capacity can be increased by the compensating condenser until it equals that from the other end of the frame to earth. The two antenna currents then become equal and produce no potential across C.

Bellini-Tosi System. The earliest direction-finding system to be generally used was the Bellini-Tosi system, in which two large loops, each as a rule composed of a single turn of wire and either rectangular or triangular in shape, are erected with their planes at right angles to each other. Two small coils are connected in series with the large loops. These two coils are arranged at right angles to each other and are placed with their planes parallel to the large loops to which they are connected. A small 'search' coil is pivoted to rotate within the two small coils, and a pointer is attached to it, arranged to move over a horizontal scale.

The two small coils reproduce, in miniature as it were, the field in which the two large loops are situated, and it can be shown that turning the search coil between the two small coils is equivalent to rotating a single loop directly in the field due to the received waves. Thus, by finding the position on the horizontal scale at which the current in the search coil is a minimum or zero, the direction of arrival of the waves exciting the large loop can be found. For accurate working it is necessary that the two large loops should be set exactly at right angles and that the loops and their circuits should be exactly similar. The same sources of error due to 'antenna effect' and pick-up by leads, &c., apply also to the Bellini-Tosi system. The circuit arrangement of a Bellini-Tosi set is shown diagrammatically in Fig. 144 (*a*). *A* and *B* are the large loops at right angles and *aa* and *bb* the corresponding field coils, while *CC* are the tuning condensers of the loop circuits. The small rotating 'search' coil is shown at *F*. The whole system of field coils and search coils is termed a Radio Goniometer.

The Robinson System. In this system, developed by J. Robinson in 1918 specially for use on aircraft, two coils *A* and *B* are used, which are fixed at right angles and can be rotated as a whole about a vertical axis. The coils are connected in series and the coil *B* can be reversed with regard to the other by means of a switch. The arrangement is shown diagrammatically in Fig. 144 (*b*). In one position of the switch the E.M.F. induced by the signal in coil *B* will be added to that in coil *A*, and in the other the E.M.F. will be subtracted. If *A*, which contains the greater number of area turns, is perpendicular to the direction of arrival of the waves, no E.M.F. will be excited in *A*. Thus when the switch is thrown over, no alteration of

signal strength will be produced, since the effect due to B will be the same in both positions of the switch. In other positions of A the signals heard in the telephones will be unequal in intensity in the two positions of the switch. Hence, if the coils are rotated until equality of signal strength is obtained in both positions of the switch, the coil A can be considered as in the

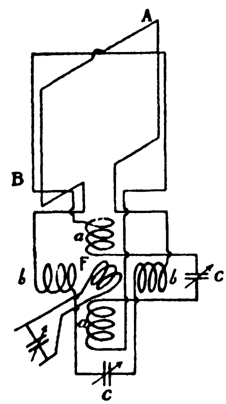


FIG. 144 (a).
Bellini-Tosi System.

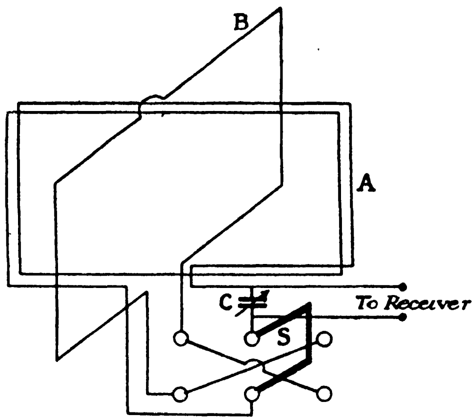


FIG. 144 (b).
Robinson System.

minimum position for reception from the transmitting station.

It can be shown mathematically that in theory the cross coils can be replaced by a single equivalent coil, and that each time the switch is reversed this equivalent single coil is displaced through an angle of, say, 2ϕ. Hence, when equality of signal strength is obtained, the equivalent coil theoretically is moved between two positions symmetrical about its zero position. The angle of 'swing', 2ϕ, of the equivalent single coil can be adjusted to the most convenient value by selecting the most suitable ratio of the area turns of the two coils A and B.

With due regard to constructional details and careful screening of the apparatus, the three systems of direction-finding above considered can be made to give results of the same order of accuracy. In a permanent station, especially on land, the Bellini-Tosi system has a certain superiority as regards robustness, quickness of operation, and the rapid detection of unreliable readings. The single-coil system has the advantages of portability and low cost, while the Robinson system is of

special value where external noise is a disturbing feature, as for example in aeroplanes.

Wireless Navigation and the Radio Compass. As soon as practical methods had been developed for finding the direction of radio waves, it became clear that the navigation of ships at sea could be assisted. There are times when it is impossible to see the sun, the stars, or any landmark whatsoever. This would be the case in fog, when the well-established navigational methods cannot be applied. It is at such times that the dangers of sea and air transport are greatest. Wireless direction-finding provided a new method for finding the bearings of fixed points and thus of establishing the ship's actual position, and one which can be used under all weather conditions.

The first method of direction-finding which proved sufficiently practical for general marine use was the Bellini-Tosi, the principle of which we have already explained. This began to come into use about 1914, during the opening years of the Great War. Previous to this the loop system had been little used, principally on account of its limited sensitivity to weak signals and its tendency to night errors. The larger Bellini-Tosi aerials could be permanently installed on board ship, and calibrated once and for all to allow for permanent errors due to the ship's hull and metalwork, whilst the small goniometer coils could be situated in the operating cabin. Bearings could thus be taken under normal operating conditions, and without the inconvenience of a rotating aerial.

There are two distinct methods of direction-finding by which it can aid navigation, and these should be clearly distinguished from each other, as they differ very much in their applications. The first method employs a number of direction-finding stations at fixed points on shore. When a ship wishes to know its position it calls these stations and transmits suitable signals to them. As many as possible of the stations receive these and measure the bearing of the ship from themselves. They then communicate these bearings to a central control station where the bearings are plotted on a chart to find the point at which they intersect. If the bearings are accurate this point will be the position of the ship. It is to reduce the chances of error that a number of bearings by different stations are required. Any two bearings would determine its position, but this would be wrong unless both bearings were exact. By the use of a larger

number it is possible to reject those which differ markedly from the majority, and are probably the least accurate. Finally, the correct position is transmitted to the ship from the control station. This is a standard method of wireless navigation which has proved of inestimable value in the past. It demands an elaborate network of stations, however, and in consequence is giving place increasingly to the second method.

In this form of navigation direction-finding equipment is installed on the ship itself, when the ship's officers can take bearings on any fixed stations that may be transmitting at the time. This method was not introduced quite so early as the former because it was found difficult to overcome errors due to the metal hull of the ship, but once this difficulty had been overcome it became standard practice. The economy of the system will be obvious since a single direction-finding equipment replaces a network of land stations. In addition the ship is able to determine its own position even when in waters remote from any organized system of land stations, provided that it is in range of any transmitters whatever whose position on the chart is known. To-day this will be the case in almost any part of the globe. Navigation in important areas is assisted, however, by the provision of special shore stations which send out continual signals specially for ships to determine their bearings. These are termed radio beacon stations.

There are a number of radio beacons in use which have been ingeniously designed so that a ship not fitted with direction-finding equipment can determine its bearing as viewed from the beacon. These beacons are fitted with directional aerials which send out a beam of radiation, such as a large loop. They are continuously rotated at a low speed, and emit distinctive signals which correspond to the direction of transmission at any instant. A ship fitted with an ordinary receiver listens to the beacon signals, and observes their nature or the time at which they pass through zero intensity. From these observations the operator can tell his direction relative to the beacon, since he will hear only the signals which he knows it to send out in that direction.

For the growing aerial transport services wireless navigation is even of greater importance than for ships at sea. For some time, however, it was only possible to use the first type of navigation, in which a group of direction-finding stations on the

ground is employed. This is used to-day on many of the air lines, direction-finding equipment of the Bellini-Tosi type being found at most of the principal aerodromes. For military purposes particularly, however, it would be preferable to use the second method, but it was found very difficult to install and work direction-finding equipment upon aircraft. The earliest methods tried out during the Great War relied on the rotation of a loop or search coil until zero signals were obtained. Owing to the noise and vibration during flight it was not possible to hear this zero accurately, and it would appear 'broad'. The search coil was therefore moved on each side of zero until the signals just became audible, when the true bearing was given by bisecting the angle between these two positions.

It was of course well known that the more usual zero position gave much more accurate bearings than could be obtained by setting to the position of maximum signal strength, but under flying conditions there was the drawback that the 'width of zero' depended upon signal strength. For strong signals it might only be necessary to swing the search coil 2 or 3 degrees before signals could be heard above the noise of the engine, but for the weak signals frequently met with this angle might become 40 to 60 degrees. Bearings would then be far from accurate. Thus it was difficult to apply conventional zero methods to aircraft, when in addition to noise there was the interference set up by the ignition system of the engine, and in many cases pilots were forced to rely upon a less accurate maximum setting.

The Robinson system was evolved primarily to overcome these difficulties, and we have seen that it enables an accurate bearing to be reached whilst still listening to signals of good strength. Robinson was the first to point out that navigation of aircraft need not follow slavishly the methods evolved for marine use. In the latter case it is necessary to know an exact position and a number of bearings become necessary. This will of course often be of value to aircraft also, but conditions in the air may be very different from those at sea, and it may be sufficient to determine merely the direction in which the plane is flying. In most cases the crew of an aircraft has up to the present been small. This was the case during the War, and whilst the larger passenger liners of to-day may carry a whole time radio operator, there will still be a large number of cases such as single-seater fighters in which all duties devolve upon the pilot alone.

He will have many things to do and may be unable to spare the time to make a complete determination of his position, which, on account of the high speed of aircraft may be changing rapidly. It will thus often be sufficient for the aerial navigator to know simply that he is on his correct course.

For this purpose Robinson introduced the simplified form of direction-finding known as homing. In this case a station at the plane's destination transmits periodically and the pilot merely takes its bearing and adjusts his course along it. In one form of his apparatus the directional loops were installed round the wings or body of the aeroplane, and the pilot turned the whole aeroplane until his homing bearing was correct, when his course was then automatically correct also. This type of navigation was of great value during the War, and is still one of the greatest safeguards against accident in the air. There can be little doubt that its more general adoption to-day would have saved a number of lives. Aerial navigation has now become a very important art so that the various modifications that have grown out of the Robinson homing system have come to be known as the radio compass.

This section would not be complete without a reference to the several forms of blind landing systems which are undergoing development to-day. These are primarily intended to enable a pilot to land his plane on an aerodrome in fog or when visibility is poor. They depend upon a combination of directional ultra-short wave beam transmitters installed round the aerodrome, and both plane and aerodrome must be fitted with the corresponding system before any use can be made of it. In the first place, the plane will be guided upon its correct course towards the aerodrome by some form of homing system of the kind described. Modern forms of this give a visual indication of deviation to the right or left of a correct course by an indicating instrument upon the dashboard. Upon reaching a certain distance from his destination the pilot flies into the field of a vertical radio beam, resembling a wall of radiation, and which warns him to begin his descent from the correct height at which he will have been flying according to his altimeter. A second similar beam may be used to tell him when he passes over the boundary of the landing-ground, or when it is time to flatten out, each beam operating a simple visual indicator such as a red light upon the instrument-board. An inclined radio beam

may be used down which the pilot can fly at the correct landing angle, and any deviation from which is at once shown by his instruments. By such methods landing in fog has been rendered comparatively safe. There is little doubt that a standardized form of the equipment will eventually be found at all major airports, but development has not yet reached that stage. It seems probable that the application of radio to aircraft instruments and navigation is still in its infancy, and a wide range of applications are undergoing development at the present time.

Errors. Apart from the instrumental errors already referred to, direction-finding installations of all systems are liable to certain errors. These may be divided into two classes: (1) permanent errors, (2) variable errors. In general, any local effect tending to distort by electrical disturbances the received waves, as they approach the direction-finder, will introduce errors in the observations. For example, if another aerial near the direction-finding loop is tuned to the same incoming wave, the direction-finder will be acted upon both by the field due to the incoming wave and by a secondary field due either to induction or to re-radiation from the local aerial. The resultant magnetic field, parallel to which the loop of the direction-finder is placed in the zero position, will be combined from the magnetic forces of the two waves, and will no longer represent the true direction of arrival of the waves from the distant station. A medium-sized aerial, about 100 yards from the direction-finder and tuned to the same wavelength, may for this reason introduce a permanent error of 4° or 5°. This error is reduced on detuning the neighbouring aerial. The presence of a tall tree near the direction-finder has a similar effect, since the tree will act practically as an untuned vertical aerial. Any large number of trees may introduce a serious permanent error, amounting perhaps to 10°, and for accurate working it is necessary to be at least from 100 to 200 yards from clumps of trees. Similarly it is necessary to avoid overhead telegraph wires, or power or lighting cables. In the same way, sheets of metal near the receiver will have currents set up in them by the incoming waves which will produce various degrees of error in the apparent bearings obtained.

In particular, the metal hull of a ship produces errors in the reading of a direction-finding apparatus installed aboard. The error in the observed bearing of a station varies, in this case,

with the direction of the incoming waves relative to the ship's axis. The rigging, funnels, hatch covers, &c., contribute largely to the error, and unfortunately the effect of some of these disturbing elements may change with the state of the weather. After every care has been taken to arrange the apparatus in the most suitable condition a correction curve is plotted for the ship by taking observations on stations in known directions.

When the Bellini-Tosi system is applied on board ship, it is a frequent practice to arrange one of the loops by adjusting its constants so as to compensate for the ship errors. When this is done, the correction curve is unnecessary.

The electrical conductivity of sea-water is very much greater than that of dry land, and, on the analogy of optics, a refraction effect is to be expected when the electromagnetic waves cross the boundary between surfaces of different conductivity. This effect is only perceptible with waves below about 1,000 metres in length. Waves below this limit, when passing obliquely from a material of higher to a material of lower conductivity, are refracted. This effect produces an error in the observed bearings in the case of signals passing along a coast-line. The error from this cause is as a rule only a few degrees and is fairly constant. When a shore station is carefully calibrated by carrying out observations on stations in known positions, this type of error can be allowed for, and the necessary correction made in the apparent bearings observed.

In the case of a land station, errors such as the above, which are due to the situation of the receiving station, can be reduced, by careful selection of the site, to 2° or 3°. A long series of observations carried out under the direction of the Radio Research Board has shown that an accuracy of this order, which is sufficient as a rule for all practical purposes, can be consistently obtained in observations upon transmitting stations which are situated only a short distance from the receiving station. There are, however, other errors which are variable and are not due to the location of the station or to the instruments. When, for example, the distance between the stations is increased, large and sudden variations of bearing occur during the dark hours.

The minimum distance at which these variations occur is roughly 15 miles when transmission takes place entirely over land, and about 90 miles when transmission is entirely over sea. The first indication that the distance between the stations is

such that variable errors are likely to occur is that the sharp, well-defined minima obtained at short distances become flattened and blurred, so that accurate fixing of the minimum position is difficult. When the distance between the receiving and transmitting stations is still further increased, the observed position of the minimum of signal intensity is frequently found to be many degrees from what is known to be its correct value. Errors of this nature are liable to begin near sunset, and continue to arise erratically throughout the night until a short time after sunrise. On some occasions the rate of the variations may be fairly rapid, amounting to perhaps several degrees per minute, at other times the variations may take the form of a slow drift of the apparent minimum.

For a number of years these 'night errors' and similar vagaries of direction-finding equipment remained unexplained, but with the discovery of the ionized layers described in the preceding chapter it was realized that indirect propagation could be held responsible for the majority of them. Bearings taken upon a station within the range of the surface wave, or during day-time and with long wavelengths which do not then experience refraction, are found to be reliable. After dark the medium-wave signals reach the frame aerials by two paths, the ground wave and the reflected waves, the latter generally arriving at an angle to the horizontal and from a direction not necessarily that of the ground wave. Short-wave signals which arrive at distant points by indirect paths only may bear very little relation to the true geographical direction of the transmitter. They may even arrive from several directions simultaneously, having no precise direction of origin. It is to interference between direct and indirect waves that most errors in direction-finding equipment are now known to be due.

The simple frame aerial, and the Bellini-Tosi and Robinson systems derived from it, are very susceptible to these errors. The most usual case of combined surface and reflected waves is that in which the latter arrives in a downward direction, as if emanating from a point in the heavens well above the horizon, corresponding to the region at which reflection from an ionized layer has occurred. This will lie approximately in the true direction of the station, and of course the strength of the reflected wave will be subject to fading. Now a frame aerial

rotating in a horizontal plane possesses top and bottom arms which are always horizontal. A wave arriving horizontally, such as a normal surface wave, will not induce currents in these arms, the signal depending only upon the phase difference between those currents induced into the vertical sides of the frame as already explained. A downcoming wave making an angle Δ with the horizontal can be resolved into components, namely, a horizontal component of intensity $A \cos \Delta$, and a vertical component $A \sin \Delta$, where A represents the momentary amplitude of the downcoming wave. The first of these components will combine with the surface wave arriving horizontally, and will either add or subtract from its amplitude according to the phase relationship at any instant. We have seen that this will give rise to fading effects, but if the direction of the two waves be the same, will not prevent a true bearing. Should the reflected wave arrive from a different or variable direction, no complete minimum direction for the frame can be found, and precise directivity becomes impossible.

The second component of the downcoming wave will behave as if emanating from a point vertically above the receiving aerial. It will thus induce no current in the vertical arms of the frame, but will set up currents in the top and bottom which are unaffected by rotation. These will form a component of signal that remains unchanged as a bearing is sought, and will not allow any complete extinction of the incoming signals for any position of the frame. Thus it is clear that accurate direction-finding with such an aerial is only possible in the absence of downcoming, scattered, or irregularly polarized waves, which means that for short waves outside the range of the surface wave it is practically useless. A decided improvement is to be expected if the aerial system could be made responsive only to waves propagated parallel to the surface of the earth, whilst being entirely insensitive to those propagated vertically.

Confirmation of this explanation has been given by the results obtained from tilted frames. On occasions when bearings on medium-wave stations have become blurred after dark, it has sometimes been found that if the frame aerial be inclined to the vertical and rotated about a tilted axis, thus placing the sides perpendicular to the supposed direction of propagation of a downcoming wave, then sharp bearings have once more been obtained. This arrangement is not very convenient practically,

since the angle of tilt must be determined experimentally for each station received.

The Adcock directional system has been evolved more recently to overcome these defects, and to make direction-finding of short-wave transmitters more accurate. It is basically a reversion to the use of two independent vertical aerials, as shown in Fig. 139 at the opening of this chapter. No connexion between their top ends to form a loop is used, and the two aerials are joined at the lower end by a screened conductor which cannot pick up signals. We saw that the phase difference between two such vertical aerials would depend upon the direction of waves arriving horizontally, whilst no current will be induced into either of them by any vertical component forming part of a downcoming wave. At the centre of the screened bottom conductor is placed a coupling transformer, corresponding to L in Fig. 139, from which a potential corresponding to the difference between A and B is passed on to the screened receiver.

The two vertical aerials of an Adcock direction-finder will be formed of rods relatively short compared to the wavelength, and supported at each end of a horizontal beam or structure so that the whole can be rotated about a vertical axis midway between them. Alternatively they may be fixed, and coupled to a pair of field coils and a search coil as in the Bellini-Tosi system. The success of the Adcock system is largely dependent upon the methods adopted to prevent any reception of signals by the horizontal leads which connect each aerial to the coupling transformer L. One method is that mentioned in which these leads are enclosed in an earthed metal tube, sometimes buried in the ground. At very short wavelengths it is difficult to earth this tube perfectly, as if it be at earth potential at one point only, signal currents may be induced into the tube itself and thence into the inner conductors. Later in this chapter we shall study non-radiating (and therefore non-receptive) transmission lines by which energy can be transferred from an aerial to a load circuit. This method is well suited to the present case, the two vertical aerials being each broken at their mid-points and coupled by such screened or fieldless lines to the load L. Fig. 145 shows a similar arrangement which will help to make the construction clear.

As a result of the elimination of all response to vertical com-

ponents of radiation the Adcock aerial is considerably more accurate and will give sharp bearings under conditions in which a frame aerial would be useless. It is thus superseding the latter for marine and aviation purposes, and all cases except those in which extreme portability is essential. As an example, a typical 'night error' of 20° to 30° when using a frame may be reduced to 2° or 3° with the Adcock system. Bearings will be obtainable from distant short-wave stations, but these may not always be strictly correct, indicating not so much the direction of the station as of the area from which the waves are reflected. At shorter ranges, however, variable errors are largely eliminated, at the expense of a bulkier aerial system and increased amplification.

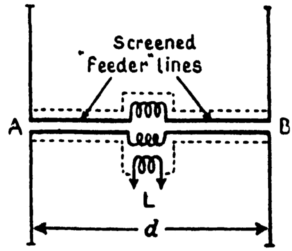


FIG. 145. Adcock System.

Resonant Aerials. The earthed or Marconi aerial and the several directional receiving-systems have now been outlined, the former being used when long waves or signals from all directions are required, and the latter when it is desired to determine the direction of incoming waves. These are the most usual receiving cases. The development of short-wave transmission has led also to the evolution of aerial systems suited particularly to transmission, which either favour a particular wavelength or concentrate radiation into a selected direction. These are termed 'resonant' and 'beam' aerials respectively, and play a very important part in modern radio-communication. To understand these we must revert to the fundamental conceptions of Chapter III. The distribution of voltage and current in a simple vertical aerial is shown in Fig. 146 (*a*), where the full line represents the current and the dotted line the voltage. The natural wavelength is four times the vertical height of the aerial.

The distribution of voltage when a coil is introduced to tune the aerial is shown in Fig. 146 (*b*), which indicates the rapid rise of potential across the coil. If, on the other hand, a condenser is introduced, the natural wavelength of the aerial is shortened, and, since the potentials of the two plates of the condenser have opposite signs, a second point X of zero potential occurs higher up the aerial, as shown in Fig. 146 (*c*). If a long aerial is excited by a short wave, stationary waves are set up in the aerial, as

in the case of the long horizontal wires considered on p. 443, and the distribution of voltage is similar to that shown in Fig. 146 (*d*). The portion of the aerial between *A* and *B* radiates as a complete Hertzian oscillator. The same effect may be produced with a shorter aerial by the introduction of a series condenser, which can be adjusted so that the upper part of the aerial has

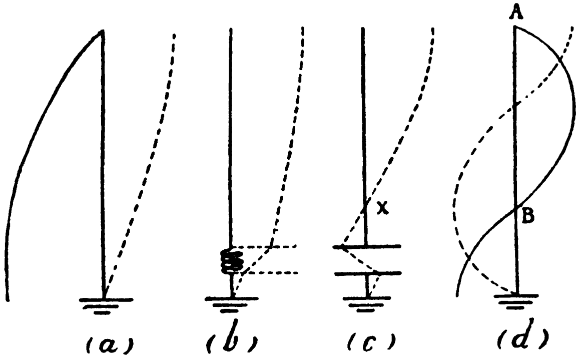


FIG. 146.

a distribution of potential and current similar to Fig. 146 (*d*).

With very short waves a method of measurement is employed depending on the production of stationary electric waves in a straight long wire. If one end of a long rope is attached to a wall and the free end held in the hand so that the rope is almost taut, then if the rope is moved up and down it is well known that waves, which appear to be standing still, are produced in the rope, as shown in Fig. 147 (*a*). The succession of points marked *V V V* in the rope remain at rest, while the points marked *III* oscillate up and down between the amplitudes given by the full and dotted curves. This phenomenon is produced as follows. When the free end of the rope is moved, waves are propagated along the rope and are reflected back at the point at which it is attached to the wall. The original waves and the reflected waves interfere with each other to form an apparently stationary wave. Obviously the point at which the rope is attached to the wall cannot move and must be one of the points we have called *V*. An analogous phenomenon can be produced by electrical waves excited in a wire ending in an insulator, the electrical waves being reflected back at the insulator. In the electrical wave the points *V V V* are those at which there are high potentials

but no current, while the points *III* are those at which there is zero potential and maximum oscillation of current. If any point of the wire is attached to earth, this point must be one at which there is zero voltage and maximum current, while the

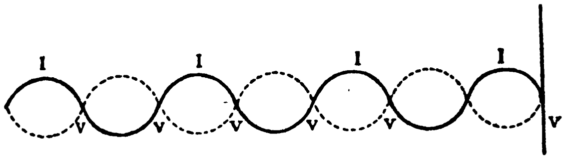


FIG. 147 (*a*).

end of the wire is a point of high potential oscillation and zero current.

In order to use these stationary or standing electric waves for measuring wavelengths, the oscillator whose emission it is desired to measure is coupled inductively to a pair of such wires. A small exhausted glass tube containing a little helium or hydrogen is then connected by small metal hooks, the ends of which run through the glass, between the wires, as shown. As

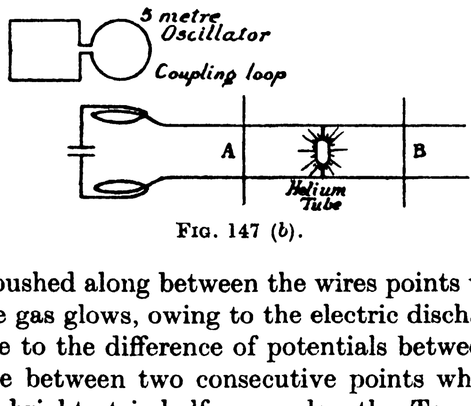


FIG. 147 (*b*).

the tube is pushed along between the wires points will be found at which the gas glows, owing to the electric discharge through the tube due to the difference of potentials between the wires. The distance between two consecutive points where the tube glows at its brightest is half a wavelength. To get the wavelength more accurately, the tube is placed at one of the points where it glows brightest and a wire *A* is placed so as to short-circuit the two long wires. If *A* is moved to a position such that the tube glows equally well whether *A* is across the wire or not, then at this position *A* will be joining two points which are at the same potential, i.e. two points corresponding to *V* in Fig. 147 (*a*). If, then, a similar position is found for a piece of wire *B* on the other side of the tube, the distance between *A* and *B*

must be equal to the distance between two of the points VV and is therefore half a wavelength. This distance can be measured with a metre scale and the wave emitted can then also be determined.

The Hertzian dipole illustrated in Fig. 19 is the prototype of all aerial systems. It is the fundamental open oscillatory circuit, consisting of two capacities joined by a (theoretically) pure inductance. From p. 47 onwards we studied the Marconi aerial, which consists of a dipole in which one half comprises a long wire having distributed inductance and capacity, completed by its 'electrical image' buried in the earth. The proportion of power radiated from this aerial was seen to be measured by its radiation resistance, a factor given by $1{,}600(\alpha h/\lambda)^2$. Neglecting the constants α and 1,600, which depend upon the design of the aerial and the units chosen, it is clear that radiation increases as the aerial height or length h increases, and also as wavelength is reduced. This gives at once a clue to the increased performance of short-wave transmitters, in which a far larger proportion of the aerial power can be usefully radiated.

The expression for radiation resistance continues to increase with these factors until a point is reached at which $h = \lambda/2$, when resonance to the transmitted wavelength occurs and efficiency is a maximum. When long waves only were used it was physically impossible to employ an aerial of this great length, but with the steady reduction of wavelength that difficulty disappeared. Moreover, it was no longer necessary to double the effective length by using an earth connexion, for a half wavelength of wire became of manageable dimensions. The short-wave aerial evolved from a Hertzian dipole thus becomes a single straight length of wire, and if this be assumed to be suspended in free space far removed from the earth, it will resonate when the total length of wire $h = \lambda/2$. Suspended near to the earth, its resonant length becomes somewhat less than $\lambda/2$. This means that the distribution of current or potential along the wire is as shown in Fig. 147 (c). The oscillatory current is a maximum at the centre, being clearly zero at each end if perfectly insulated. The potential is oppositely distributed, being very low at the centre and a maximum at each end, as between A and B of Fig. 146 (d).

We shall call this aerial a 'half-wave' aerial, and it is the 'unit' from which beam arrays and other complex systems are

built up. Frequently this half-wave unit is called a dipole, or doublet. Neither of these terms is strictly applicable, the dipole being two capacities, joined by pure inductance, whilst the doublet implies a pair of electrical charges. Since the half-wave aerial has no direct connexion with the earth, it may be suspended either vertically or horizontally, the latter being pre-

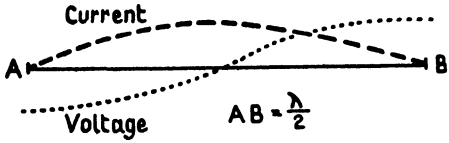


FIG. 147 (c).

ferable in most practical cases because as has been mentioned the shorter wave signals tend to become horizontally polarized after ionic refraction. Radiation from the aerial in free space is similar to that of the Hertzian dipole, the direction of propagation being in a plane perpendicular to the length of the aerial. The electric field will lie in a plane parallel to the aerial wire, whilst the magnetic field will be mutually perpendicular to the wire and to the direction of propagation. We thus see that the half-wave aerial is inherently directional, and if in the horizontal plane, will transmit best towards certain points of the compass.

Since both oscillatory current and voltage vary along the half-wave radiator, it will naturally represent a differing impedance along its length, as viewed from any outside circuit to which it might be joined. If remote from the ground, this impedance is independent of frequency for any aerial oscillating at its resonant frequency. It will be about 73 ohms at the centre of the half-wave where current is a maximum, and will rise uniformly to about 2,400 ohms at each end. These facts will be found very important later when we come to consider how power is fed to complex aerial systems.

The following figures will give some idea of the size of half-wave resonant aerials for various frequencies. They are all examples that can be worked out from the general formula for aerial length, which is:

$$\text{length of half-wave resonator} = 0{\cdot}475\lambda = \frac{467{,}400}{F} \text{ feet,}$$

where λ = wavelength in metres, or F = frequency in kilo-

cycles. They are correct for an aerial at a moderate height above the ground, and between the limits of 3,000 and 30,000 Kc., outside which the correction for 'end effect' varies slightly. For a wavelength of 320 metres the length of wire would be about 530 feet, a considerable but not unmanageable length. It is thus possible to erect steel lattice towers which themselves form resonant aerials for wavelengths in the medium broadcast band, and vary in height from some 200 up to nearly 1,000 feet. The resonant aerial does not become convenient, however, until the true short-wave region below 100 metres is reached. For an 80-metre wavelength the aerial would be some 125 feet in length, for 40 metres about 66 feet, whilst for 10 metres it would be only 16 feet.

In practice the wavelength at which a given length of straight conductor resonates is not entirely independent of the thickness of the wire, of the proximity of the ground, and of surrounding objects, and the resonant wavelengths will lie between 2·1 and 2·05 of the physical length. The fact that slightly less than a half-wavelength of wire is necessary arises partly because radio waves do not travel quite so quickly along a conductor as in free space, and because the aerial can never be entirely out of the influence of the ground, except perhaps when suspended from an aeroplane. A good approximation for the resonant length is $L = 0{\cdot}475\lambda$ in metres $= 1{\cdot}56\lambda$ in feet.

It is an interesting fact that the radiation resistance of a half-wave aerial in free space is approximately 70 ohms, and is not dependent upon wavelength. It is perfectly possible to excite an aerial at harmonic frequencies, or sub-multiples of the natural wavelength, just as in the case of a vibrating violin string. This is a very apt comparison, the string being clamped at each end and having maximum amplitude at the centre, just as the current nodes occur at the ends of a half-wave wire and the maximum oscillatory current at the centre. The aerial will thus oscillate when it equals any number of half-wavelengths at the frequency by which it is excited, and thus a 40-metre half-wave unit for example could be used on 20, 10, or 5 metres. Radiation resistance rises at the harmonics. If the aerial be two half-waves long it becomes about 90 ohms, at four half-waves it rises to 110 ohms, reaching double the original value at about five wavelengths. Thus, slightly more power will be radiated from an aerial when operated at a harmonic than at the funda-

mental, but it is necessary to see that the resonant length is accurately adjusted to the correct value.

We have seen that a half-wave aerial is more nearly $0{\cdot}475\lambda$ long, owing to certain modifying influences which may be termed 'end-effects'. No such effects occur in the case of the inner half-wave portions of an aerial oscillating at a harmonic; and the total length becomes more nearly $n\lambda/2$. If the aerial be used at its nth harmonic, the end correction need only be applied once, and so the expression for total length becomes $L = 0{\cdot}475\lambda+(n-1)\lambda/2$. Thus we see that an aerial will not resonate at true harmonics which are exactly n times its fundamental frequency, but at slightly higher frequencies; and if designed for harmonic operation, as may often be convenient, its length must be suitably corrected.

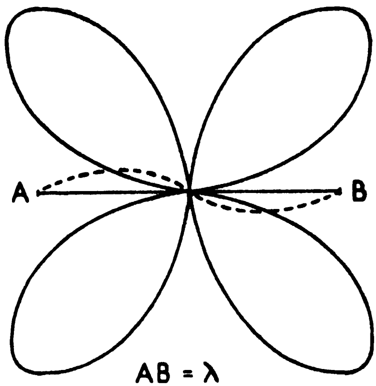


FIG. 148. Radiation from a Full-wave Aerial.

The directional properties of half-wave and harmonic aerials in free space are quite simply arrived at. The half-wave unit will radiate nothing from its ends, and the polar curve will be exactly similar to that for a frame aerial shown in Fig. 141, in which the line YY' represents the aerial wire. An aerial two half-waves long, also termed a 'full-wave' or harmonic aerial, exhibits four 'lobes' of maximum radiation in the horizontal plane, as shown in Fig. 148. For a four half-wave aerial the number of lobes increases to eight, four being major and four minor; whilst at eight half-waves there will be 16 lobes of radiation. At any considerable distance from the aerial, however, these more complex patterns cease to convey much meaning, for the effects of refraction and scattering of the radiation tends to even them out. Thus the signal strength from a one or two half-wave radiator will show pronounced directional effects, particularly within the range of the surface wave, but the higher harmonic aerials tend towards uniform radiation with actually a predominance towards the ends of the wire where the major lobes coalesce.

Used vertically, the resonant half-wave aerial will clearly tend to radiate equally in all directions in a horizontal plane. In practice, however, this fact is modified by the presence of the semi-conducting earth, which reflects a large proportion of the radiation reaching it as shown in Fig. 149. As a result radiation is concentrated at an angle θ to the horizon depending upon

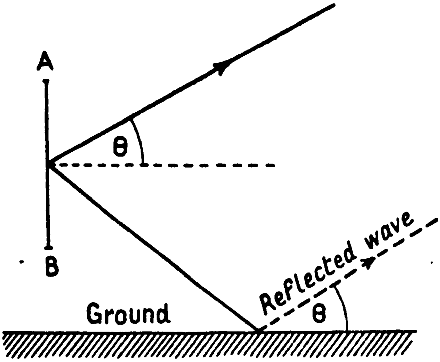


FIG. 149. 'Earth Reflection'.

the height of the aerial above ground, no radiation taking place theoretically parallel to the ground. The factor of 'angle of radiation' above the horizon is clearly an important one in distant short-wave communication by the indirect waves, since it affects the angle at which most of the radiation strikes the ionized layers of the upper atmosphere. A low angle of radiation is found most effective for the longest ranges of from 3,000 up to 10,000 miles, and aerial systems are designed to radiate as far as possible between 5 and 20 degrees to the horizontal. This condition is favoured when a vertical half-wave aerial is erected with its lower end about a quarter-wavelength above the surface of the ground. This is taken, however, as the effective electrical surface, and is best determined experimentally.

Earth reflection has a pronounced effect also upon the effective directional characteristics of horizontal aerials. These radiate in accordance with the polar curves given if measured within the range of the surface wave, but their actual range at a distance obeys quite different laws. This is due once again to angle of radiation above the horizon. The horizontal half-wave aerial, radiating in all directions perpendicular to the wire, will tend to radiate much of its power nearly vertically. Reflection

by the ground of the downward radiation will tend to concentrate this at high angles, which will be reflected by ionized layers to return to the earth at limited ranges only. There will be little radiation once again parallel to the ground, owing to absorption by trees or buildings and the like. To increase this and to reduce the earth losses a horizontal aerial should be suspended as high as possible, the practical minimum being about a quarter-wavelength.

Harmonic aerials are also affected by reflection from the earth in a similar manner, and the effective polar curve for long range communication will differ from those measured locally. The precise effect will depend upon local ground conductivity and surrounding objects, and is thus difficult to calculate, being determined experimentally for individual aerials. In general, long aerials excited at their harmonics produce lower angle radiation than the half-wave unit, and this is more uniformly distributed in the horizontal plane, with the result that they are effective for all round long-distance working.

To sum up the foregoing paragraphs, it can be said that the resonant or Hertzian aerial is more efficient as a radiator than the earthed or Marconi aerial, and is therefore universally used when the wavelength is short enough to make it practicable. The half-wave unit is the shortest resonant length, and possesses pronounced directional properties which are modified in practice by earth reflection. Harmonic aerials of greater length possess somewhat greater radiation resistance and efficiency, and are less strongly directional in practical surroundings. Radiation should be concentrated at low angles to the horizon for long distance signalling by the reflected indirect waves, say between zero and 15 degrees, but angles more nearly 45 degrees are suitable to shorter distances just beyond the skip distance. To obtain low-angle radiation vertical aerials having their lower ends from a quarter- to a half-wavelength above ground can be used, or a horizontal aerial must be as high as possible above the earth.

Sloping aerials will possess modified polar curves, favouring low-angle radiation from the lower end and little from the upper. Clearly there are a very wide range of possibilities which cannot be fully discussed, but they are of secondary importance because the directivity of simple resonant aerials is not extreme. It is masked at long ranges by scattering, diffraction

and reflection of the waves, which may thus reach regions not expected from the local polar curves. More complex directional systems are used in practical working. No mention has been made of the connexion between resonant aerials and the transmitter, which will be fully explained later in the chapter, and which employs a transmission line. For the present the aerial should be regarded as fed with power by a method that has no effect upon its performance in other respects.

For broadcast transmission aerials of the types described are employed in practice, and they are also suited to general reception. It has been found, however, that the effective concentration of radiated energy into a narrow beam is an enormous advantage when communicating between fixed points, and there is an increasing tendency to use this method whenever possible in commercial installations. The loops of a polar curve such as that of Fig. 148 are termed 'lobes', a major lobe being one into which a large proportion of the radiated energy is confined. The simpler resonant aerials radiate several lobes, the power being distributed between these; but for point-to-point communication aerials are designed to concentrate power into a single major lobe which is accurately directed towards the desired receiver. By definition of the polar curve, the radius in any direction measures the field strength in that direction, and the gain in strength will thus be proportional to the length of a lobe. Fig. 150 illustrates the kind of polar curve obtained from a good directional array, and it can be shown that not only is the field strength proportional to the length of the lobe, but the power radiated in that direction is proportional to the area of the lobe. A circle would represent the radiation from an ideal vertical aerial excited with the same power, and so the ratio (radius of circle to length of lobe) measures the 'gain' of the array, usually expressed in decibels.

It will be quite obvious that if an aerial system radiates several times the power in a desired direction, it will be possible to reduce the power of a transmitter in the same proportion whilst retaining an unchanged field strength at the receiver. This fact greatly reduces the cost of transmitting stations, so that tens or hundreds of kilowatts are now used with improved reliability over distances that would require thousands of kilowatts in a non-directional aerial system. Similar arrays can be used at the receiver, with a large gain in strength of the wanted

signals over others, and over atmospherics and local noise. The gain in the major lobes of the simpler resonant aerials will seldom exceed 4 decibels. For an efficient array it may lie between 10 and 20 decibels. Higher figures are possible, but there is a limit to the useful maximum because with increasing gain the polar curve becomes increasingly narrow and direc-

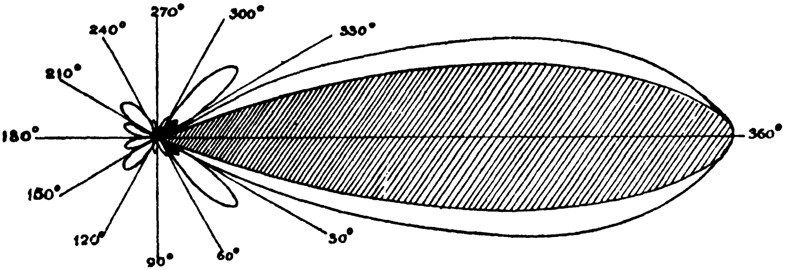


FIG. 150. Polar Curve for 15 Aerials.

Shaded portions represent energy radiated in given directions. The unshaded portions represent radiated field. Secondary maxima are shown on the left of the diagram.

tional. The apparent direction in which indirect waves must be transmitted may vary slightly with conditions in the upper atmosphere, and a certain width of beam is necessary to accommodate these changes in the apparent direction of transmission, and angle of radiation. Difficulties of design also arise, limiting the gain of a practical array to something between 20 and 35 decibels as an extreme maximum.

Aerial Arrays. There are three broad methods by which aerial systems can be made highly directional, all of which are based upon combinations of the simple Hertzian half-wave unit aerial. In his classical experiments Hertz employed very short waves excited in the oscillator described in p. 2 by means of an induction coil. He showed that the radiation from his oscillator could be concentrated into a beam by reflection from a metallic mirror. To do this he placed the transmitting aerial in the focal line of a metal sheet bent into parabolic form, and obtained sparks across a receiving oscillator placed in the focal line of a similar mirror some distance away. These sparks occurred only when the two reflectors were facing one another. Modern methods of beam transmission are based on the experiments of Hertz. In the first type groups of half-wave radiators are excited

in phase, being so placed in relation to each other that their fields combine in the desired direction whilst cancelling in others. Secondly, reflectors are still used, although made up of resonant wires rather than sheet metal, and may be combined with the grouped radiators of the first method. Thirdly, there is the inherent directional effects of long harmonic aerials, which can be increased and harnessed in several ways. We shall now consider each of these methods in greater detail.

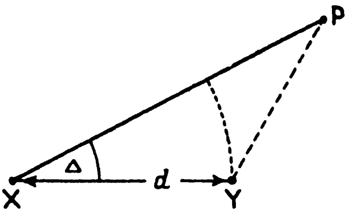


FIG. 151.

Consider two half-wave radiators placed parallel to each other, at a distance d apart. These are represented in end section by X and Y in Fig. 151. Then at a distant point P making an angle YXP of Δ to the line joining X and Y, the field strength will be the sum of that produced by each of the two radiators. Treating each as a point source radiating in the same phase, the field at P will be the vector sum of that from each source. If the distance $XP = YP$ the two waves will arrive in phase, and will add, the field strength being a maximum. This will also be the case if these distances differ by a whole number of wavelengths. Should they differ by one or any odd number of half-wavelengths, however, the two waves will be in opposite phase, and the resultant will tend to zero.

The separation d between the radiators will be a deciding factor in the combined polar curve. When this is small compared to the wavelength, the two radiators will behave very much like a single aerial, the polar curve being approximately circular. This will become increasingly elliptical as d is increased, the major axis being perpendicular to XY, will develop a 'waist', and when d reaches the critical value of a half-wavelength the curve will be hour-glass shaped. It will resemble Fig. 152, the lobes somewhat elliptical instead of circular. There will be no radiation along the line joining the two radiators. This can be easily seen because XP and YP will now be in the same straight line, and must therefore differ by the length XY. This has been made a half-wavelength, and so the condition for complete interference exists at all points along that line, in either direction. Equally simply it can be seen that all points along the

line AB, perpendicular to that joining the two radiators, will be equidistant from each. XP must then equal YP, and the two waves are in phase at P whatever the distance XP may be. No matter what the spacing d between the radiators, therefore, radiation will be a maximum in the 'broadside' direction, and for this reason the combination is a simple type of 'broadside array'.

As d is further increased, more complex polar curves occur.

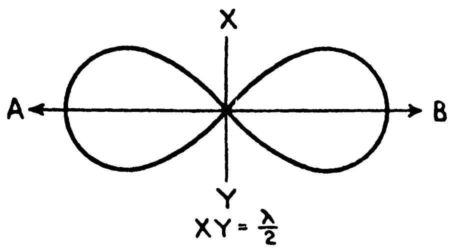


FIG. 152.

At one wavelength there will be four lobes, and at two wavelengths eight. These cases are of mainly theoretical interest only, being little used practically. The curves can be plotted by making use of the expression for the phase difference at P along any direction making the angle YXP equal to Δ. This is given by:

$$\text{phase difference} = 2\pi(d/\lambda)\cos\Delta.$$

Drawing the vector diagram, and assuming unit field strength at P from each aerial, the resultant field becomes:

$$\text{resultant field} = 2\cos\text{ (half angular phase difference)}.$$

We now have the fact that a pair of radiators excited in phase and spaced one half-wavelength (in either the horizontal or vertical plane as desired) will radiate only in the broadside direction, the radiation falling to zero as the line of the radiators is approached. Consider what will happen to the radiation if the power-output from a transmitter is shared equally between these two instead of being confined to one radiator. Let the transmitter deliver W watts to the aerials, and the mean current in a single radiator be $2I$ amperes. Then we can assume that the available power will be divided equally between the two radiators, and since in a circuit of any given impedance power varies as oscillatory current squared, the mean oscillatory current in each will be $\sqrt{2}\,I$. But the field strength at a distant

pcint due to an aerial is proportional to the aerial current, other factors remaining unchanged, not to the radiated power. Hence in the direction of maximum radiation of the pair the field strength becomes proportional to $2\sqrt{2}\,I$. There is thus an increase in field strength of $\sqrt{2}$, or approximately 1·4 times that produced by a single aerial receiving the same power from a transmitter.

This argument can readily be extended to larger numbers of radiators, on the assumption that we are concerned with field strength at a point where the fields from all are additive in the same phase, namely along the axis of a major lobe of the array. Thus four radiators, or 'elements' of the array as they are now termed, will double the field strength obtainable from a given power; whilst 16 elements will increase the field strength four fold. An increase in the actual efficiency of the aerial system as a radiator therefore takes place if a number of half-wave radiators are used 'in parallel'.

We have seen that two such elements placed parallel and separated by one half-wavelength form a simple directional array, having maximum radiation in the broadside direction. This process is extended to produce a practical design by the use of a larger number of elements regularly spaced by the same distance in a single plane. These can be treated as a number of pairs further combined into pairs, and so on. The whole will thus radiate most efficiently in the broadside direction, but the directivity becomes greater, field strength falling off much more rapidly with increasing angle from the maximum direction as the number of elements forming the array is increased. The polar curve resembles that of Fig. 150, except that there will be two such lobes radiated from the two sides of the array. The very small secondary lobes shown are unavoidable, but usually have very little practical importance.

The construction described results in the radiation of a beam over a limited horizontal angle, which can be reduced to any extent required in practice if sufficient elements be used. If the elements be vertical, radiation will also fall off with increasing angle above the horizon, becoming zero vertically upwards. There will be a rather indefinite concentration at some medium angle, but it will be advantageous to improve this also, reaping the benefit of vertical as well as horizontal directivity and concentration. To do this a second or third tier of radiators is

employed above the first, making up a 'sheet' of aerials, which constitutes a beam array in either plane. The beam will not be parallel to the ground, however, but will be at a definite low angle to the horizon owing to the effects of reflection at the earth's surface, as was explained in Fig. 149. In practical construction it would be too complex to feed every element by transmission lines of the type to be described later, and so a proportion of them are so fed whilst the remainder derive their energy by direct connexion to or coupling through short feeders from the driven elements.

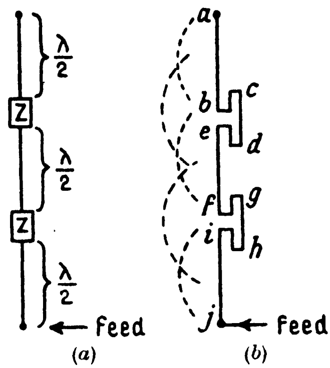


FIG. 153. Tiered Arrays.

The idea of tiered radiators can be usefully applied to a group of vertical elements, which are placed in line one above the other, and excited in the same phase as before. This aerial will have about the same length when built up from n half-wave elements as would a harmonic aerial working at its nth harmonic, but whereas in the latter case each successive element would be oscillating in opposite phase to its neighbours, in the tiered vertical array the oscillatory currents will be in the same phase throughout. Low-angle radiation tends to cancel out between the successive half-wave sections composing a harmonic aerial, and as a result the bulk of radiation takes place at large angles to the wire. The rearrangement of all elements in the same phase, however, prevents this cancellation, with the result that considerable power is radiated at low angles.

Fig. 153 (a) shows a number of vertical half-wave elements forming a tiered array. One method due to Franklin, whereby these can be excited in phase, is to insert between each a reactance that will just cause phase reversal at the frequency employed. Since the phase at the end of each element is about to reverse as energy passes along the aerial from the end at which it is excited, we can regard this reactance as a non-radiating half-wave element in which the next phase reversal takes place, so that the radiating element which follows is in the correct phase once again. The dotted curve of current distribution

illustrates this condition. The phase-reversing reactances can consist of condensers or inductance coils of appropriate value, or even of circuits resonant to the transmitted frequency. Since a reversal of phase occurs between the two ends of a parallel tuned circuit, it will be equivalent to a half-wave radiator in that respect. The array as a whole can be fed with power from one end, and although there will be a certain loss of energy before the furthermost elements are reached, due to losses and reflection at the preceding elements, the aerial as a whole will be a much more efficient low-angle radiator than a single wire of the same length. This improvement results because the several half-wave elements are in phase and the radiation from each, therefore, assists that from the others.

The ideal radiator of any given length would be a vertical wire having the same oscillatory current throughout its length. This condition can be approached by an arrangement similar to that just described, in which the phase reversing elements consist of half-wave lengths of wire bent back upon themselves in the manner shown in Fig. 153 (*b*) at the right. The aerial can now be regarded as a row of overlapping elements, *a–b*, *c–d*, *e–f*, *g–h*, &c., and it will be seen that the central portion of each where current is a maximum coincides with the end portions of the two adjacent, where current is falling off. Since the radiation at any unit length along an aerial is proportional to the current in it, it will be seen that current and radiation are now approximately uniform along the length of the aerial, and equal to the maximum of any one element. The most efficient use possible is thus being made of the available length. This construction is that underlying the Franklin uniform-current aerial, which in the vertical form is much used commercially when low-angle radiation is desired, and to minimize unwanted space waves which may interfere with reception. Placed side by side at half-way intervals, a number of such aerials make up a broadside array having directivity in both planes, very much as described in the preceding paragraphs.

It is not proposed to describe the practical construction of any particular forms of array, since these are very numerous and differ in practical rather than basic principles. All are composed of a number of half-wave elements excited in phase by whatever feeder system best suits their location and the designers' ideas. Either vertical or horizontal elements may be

used, the former being rather more general, and in either case several tiers of 'stacked' elements are preferable so that radiation shall be confined to a beam in both the horizontal and vertical planes. The number of tiered rows used will be determined by practical considerations of erection, since, except at the shortest wavelengths, it will be impracticable to erect masts of sufficient height to support more than three or four half-wave elements above each other. In the width of an array more latitude is possible, and this dimension may be spoken of as the 'aperture' of the system. It is usually of the order of 5 wavelengths, this being the best compromise between cost and efficiency. Above some 10 wavelengths the wave-front ceases to obey the simpler theory we have outlined, and the gain in efficiency falls off. Taking as an example an array of four tiers of radiators, having a 5λ aperture, it will be seen that there will be some 44 radiators, which represents about the size of a large commercial system.

So far we have described an array as a 'sheet' of aerials, radiating equally in two opposite directions, which would be perpendicular to the plane of the array in free space, but are directed upwards somewhat in practice owing to reflection from the ground. It is generally a marked improvement to suppress radiation in one of these directions, adding the power so saved to nearly double radiation in the other. This can be effected by reflectors, which form a duplicate array or 'reflector curtain' behind the radiators, and throw the back radiation forward. A reflector consists simply of an insulated length of wire, some 3 per cent. longer than a half-wave radiator element, and parallel to it at a distance of $\frac{1}{4}$ or $\frac{3}{4}$ wavelength.

The field of the radiator will induce oscillatory potentials into the reflector also, these being similar in amplitude and frequency but lagging 90 degrees in phase. The reflector will re-radiate independently, and since the potentials induced in it lag by 90 degrees, and the current which they produce by a further 180 degrees, the radiation leaves it 270 degrees later in phase than that from the radiator. In traversing the quarter-wave distance back to the radiator, this field becomes a further 90 degrees lagging in phase, making a total lag of 360 degrees, and is thus in phase with the field which produced it. The two thus combine and are radiated in the forward direction as a common field of approximately twice the original strength. In the backward

direction, that of the reflector, the radiated field is almost all absorbed in inducing current into the resonant reflector. Thus the whole effect of a reflector an odd number of quarter-wave-lengths behind a radiator can be summed up by saying that it absorbs back radiation, and subsequently re-radiates this in the correct phase to reinforce that in the desired direction.

An improvement in the reflecting process can be effected if the reflectors are fed with power in the correct phase relationship, when resistance losses can be compensated for so that the back field is exactly cancelled and the forward exactly doubled. This is not very often thought worth while in practice, however, the reflectors being generally excited parasitically as explained. Traces of back radiation then remain, but in the case of a large array having one or more reflectors behind each radiator will be negligible for most purposes. Their presence is evident in the polar curve of Fig. 150, giving rise to the small backward lobes there shown.

It is interesting to note that reflectors were the basis of the earliest serious directional experiments, which were probably those of Franklin carried out between Hendon and Birmingham in 1921. The wavelength used was 15 metres, and the distance 97 miles. The waves were directed into a beam by means of a reflector system consisting of a large number of parallel vertical wires arranged to form a parabola with the transmitting aerial at its principle focus, as shown in Fig. 154. The reflectors were tuned, and a similar reflector system was used at the receiving station. Measurements showed that about two hundred times as much energy was received when both reflectors were used as could be received without them.

The parabolic arrangement of reflectors is now little used at ordinary short waves, the flat beam arrays being found more suitable. In the ultra-short wave region, however, and particularly in the micro-wave region below one metre, it is found a convenient one to employ. A parabolic reflector similar to that illustrated is built up from short resonant copper rods or tubes, which being at most a few feet long can be supported by a wooden framework. A dipole or half-wave radiator is placed at the focus, and sometimes the whole power stage or oscillator will be of such small dimensions that it can be placed within the reflector, thus reducing problems of coupling to the radiator. Alternatively, the oscillator is placed just behind the

parabolic reflector, and a short feeder used. The whole equipment can be rotated and directed towards the receiving station, and since ultra-short waves behave very much as light waves, it is not an exaggeration to compare the whole system to a radio searchlight.

Amongst the interesting applications to which such parabolic

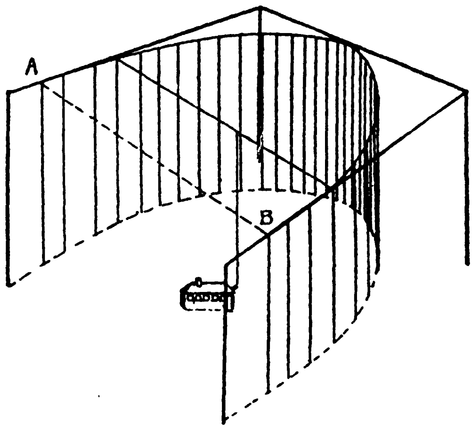


FIG. 154.

ultra-short wave beams have been applied we have noted that of the 'radio beacon' or 'lighthouse', used to inform ships at sea of their direction with respect to it. The transmitting beam is rotated at a uniform rate, which, in the case of the first installation by the Marconi Company at Inchkieth, is two minutes for a complete revolution. As the beam rotates, distinctive Morse signals are emitted corresponding to the various points of the compass. On board ship it is only necessary to observe which of these signals is the most strongly received to know the compass bearing of the lighthouse from the ship.

The third type of directional aerial system originally referred to is not built up from the usual elements of half-wave radiators and reflectoιs, but is more related to the harmonic aerials. If a long conductor containing a large number of half-wavelengths be placed end on to incoming waves, there will be a tendency for a travelling wave to be set up along it, and if the conductor be harmonically resonant at the incoming frequency this wave will reach useful proportions. A glance at the polar curves for harmonic aerials of Fig. 148 will explain this from another angle.

As higher harmonics are considered, the four major lobes of radiation (or reception) approach the line of the aerial wire. At very high harmonics these tend to coalesce and to form a cone of radiation off each end of the wire, the axis of the aerial being the principal axis of this cone. All other lobes have now become small, and so the system forms a directional aerial in an end-on

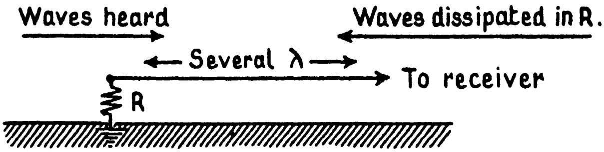


FIG. 155 (*a*).

direction. Connected to a transmitter or receiver at one end and a terminating resistance at the other, the aerial becomes unidirectional (see Fig. 155 (*a*)). Waves arriving from the terminated end build up to form a signal at the receiver, whilst those from the opposite direction build up potentials which are eventually dissipated in the resistance.

Although such a 'Wave Aerial' will respond most effectively at its harmonics, a large potential will be built up by any wave travelling parallel to the wire, provided that this is long relative to the wavelength. The system thus forms an excellent directional receiving aerial for use over a wide band of short waves, this being its principal application. In one practical form developed by the Radio Corporation of America the long wire is replaced by a parallel pair resembling a feeder line, to each of which are attached at short intervals non-resonant perpendicular collector wires. These are coupled to the long wires through condensers to prevent excess loading, and serve to increase the energy picked up by the system as a whole.

A very effective use of harmonic aerials can be attained if these are combined as a pair to form a 'V'. This is illustrated in Fig. 155 (*b*). Two harmonic aerials make an angle Δ with each other, the feeder being attached at the apex A, the whole being a plan drawing in the horizontal plane. Since each aerial is of the same length and fed from a common point, the current in them will be in phase. Let Δ be chosen to be twice the angle made by each major lobe with the line of a single harmonic aerial at the wavelength in use, then the shaded lobes will lie in a straight line, and, being in phase. will combine to give a

concentration of radiation. Those unshaded will be equal in phase and amplitude also, but opposite in direction, and will thus produce no resulting field at a distance. The 'array' must be many wavelengths long to yield its best performance, and thus takes up considerable ground space, but when this

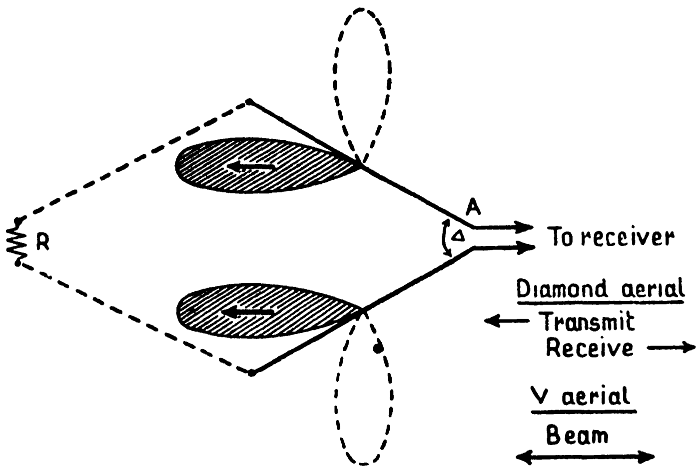


FIG. 155 (b). Long Wire Arrays.

can be provided the efficiency is equal to or better than the most elaborate broadside arrays.

If two such 'V' aerials be placed 'mouth to mouth' as shown dotted to form a diamond, employing four long harmonic wires in all, the performance is still further improved. It is now possible to terminate one apex, that opposite to the feeder, by a resistance, when the array becomes unidirectional. Under such conditions strict harmonic operation is unnecessary, and the system will work over a band of wavelengths with little alteration in performance. This occurs because non-resonant oscillations which might be reflected from end to end of the wires to form wasteful standing waves will be prevented from doing so by the terminating resistance, which will absorb them. Thus only waves travelling in the correct direction will build up potentials at the feeder, and the amplitude of these will depend simply upon the length of wire upon which they have operated, and not upon any resonance in the wires. The resistance prevents the usual resonance phenomena in each of the aerials, converting them into a unidirectional semi-aperiodic system.

We have now surveyed the principles upon which fixed directional arrays can be built up, and are in a position to appreciate the detailed designs given in more advanced works. Most of these are unsuited to direction-finding because of their great size, but it is interesting to note that by slightly modifying the phase in which successive radiators are excited, the directional properties of many types can be adjusted, the beam being 'swung' through a few degrees.

The following are a few minor points that are often overlooked, and which serve to conclude our study of directional systems. Firstly, it should be noted that however directional in the horizontal plane a beam may be, it is impossible to radiate maximum energy parallel to the earth's surface. This is explained in Fig. 149, where it is shown that radiation leaving the aerials in a downward direction will be reflected to reinforce the main beam at an ascending angle. For this reason radiation always occurs at an angle to the horizontal, truly horizontal radiation being theoretically always zero. Combined with the effects of reflection by ionized regions, this means that the effective polar curve at long distances where indirect waves only are being received will differ materially from that measured locally. This is a factor which cannot be ignored in design, an array which radiates at a vertical angle suited to the distance and wavelength being often superior to one having greater calculated horizontal concentration. It is sometimes necessary to tilt the array in order to arrive at a suitable angle of radiation, an effect of tilting being obtainable in some cases by small changes in the phases of the exciting power applied to successive elements.

Secondly, we have noted that the half-wave radiator will be actually about 94 per cent. of an actual half-wavelength long, since the rate of propagation and hence the wavelength of electro-magnetic waves is somewhat different along a conductor than in free space. The proximity of other conductors increases the resonant length, however, and in a densely built array may practically remove the 6 per cent. discrepancy. Reflectors on the other hand are not loaded by feeder connexion or the like, and will resonate at a length more nearly the true half-wavelength. All such estimates of length are influenced by the presence of the ground, and vary somewhat with height and the presence of surrounding objects. It is therefore usually

necessary to check all array designs by experiment, and to make fine adjustments after erection. Fortunately the resonance of aerial systems is not so sharp as to make this an unduly difficult process, being equivalent to that of a heavily damped resonant circuit.

Throughout this chapter we have been assuming that energy can be fed to an aerial from a transmitter, or taken from it to a receiver without modifying the performance of the aerial in any material way; and that this energy can be in any desired phase. We must now see how this can be achieved in practice.

Simple aerials such as the Marconi described in Chapters III and IV usually contain a loading coil which can be coupled magnetically to the transmitter output circuits. This coil is preferably near the earthed end of the aerial, which is near its electrical centre, since the aerial is assumed to be completed by its electrical image beneath the ground. At this point current will be large and potential low, and it is quite safe to bring the actual aerial into the transmitting building. This has always been the usual practice when using earthed aerials on long or medium wavelengths, but has the drawback that radiation from the portion within the station buildings may be absorbed by surrounding objects. It is thus lost, and may raise the loss factor of the aerial as a whole considerably. It may also cause trouble by inducing large currents in other parts of the installation. At short waves all these defects increase, and it is undesirable to bring a radiating portion of the aerial within a short distance of any buildings. The efficiency of the aerial will only be a maximum if it is entirely clear of obstructions and from a quarter- to a half-wavelength above ground, whilst in the case of arrays a clear location may be essential to the correct polar curve.

In those cases where a Hertzian or harmonic aerial is brought directly to the transmitter, it will clearly be fed from one end. This we have seen is a point of maximum potential and impedance, the reactance being infinite, and hence the aerial must be joined to a point of similar high impedance in the transmitter if energy is to be effectively transferred. A parallel tuned circuit has infinite reactance at its ends, and hence if a circuit LC resonant to the transmitted frequency is coupled inductively to the transmitter tank coil, as shown in Fig. 156, the 'end-fed' aerial can be attached to one side of this. Coupling to the

aerial can be set to the best value simply by varying the magnetic coupling between the two coils as in the case of an earthed aerial, but it must be noted that the earth plays no part in the present coupling system, an earth lead being theoretically unnecessary. It may be used in practice to stabilize the equipment, but should carry no radio-frequency current, and need not be of particularly low resistance.

An advantage of the simple end-fed arrangement is that the

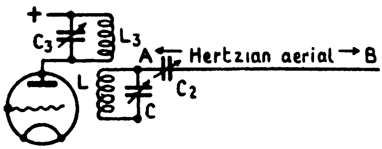


FIG. 156.

aerial can be used equally on any harmonic frequency, or can be worked at a frequency to which it is not resonant by the use of a loading coil or series condenser. Thus, if the aerial be too long, its excessive capacity can be reduced by a tuning condenser at C_2, or if too short the inductance can be brought up to a resonant value by a coil inserted at the same point. The whole system is controllable from within the station, making it suitable for marine or portable use on several wavelengths. A large point-to-point transmitter would, however, employ a separate aerial for each wavelength or service, and would prefer efficiency to the flexibility given by end feeding. The arrangement is popular with amateur and experimental transmitters.

The drawbacks enumerated above can be overcome if the portion of the aerial brought within the station buildings can be made non-radiating. The simplest method by which this can be done lies in the bending back of a portion of the aerial upon itself, so that the field of each portion will cancel that of the other. Fig. 157 illustrates this process. Let $ABCD$ represent an aerial, which we will assume to be two half-waves (a full wave) long, the point B being the mid-point. The aerial thus operates in its second harmonic, B being a point of minimum current, and both AB and BD half-wave radiating sections. The portion BCD is now bent back upon itself about its mid-point C, so that CD and BC are each a quarter-wavelength long, as in Fig. 157. Now C will be a point of maximum current, which decreases along both CB and CD to become zero at B

and D respectively. Thus at any point along these two parallel wires the oscillatory current will be equal in magnitude but opposite in phase, and the field produced by them at a distant point will thus be zero.

To utilize this scheme in practice the point C is taken into the station and is coupled to the transmitter, a small inductance being inserted at that point to allow of magnetic coupling. The length BC must be sufficient to place B clear of all obstructions,

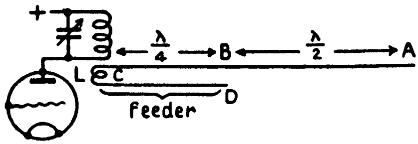


FIG. 157.

when the half-wave radiator AB can be well removed from all surrounding objects. The signals are only radiated or received from this unscreened portion, the parallel wires CB and CD being termed a 'non-radiating tuned feeder' although they are electrically part of a harmonic aerial system. The whole system is known as a 'Zeppelin aerial', as it was first extensively used on airships of that type.

It may easily happen that the length BC is insufficient at short wavelengths to bring the radiating portion of the aerial clear of obstructions. Since we are dealing with a harmonic aerial, however, there is no objection to increasing the total length to any number of half-wavelengths, when the feeder portion can be any convenient number of quarter-wavelengths in length. If desired, the radiator may also comprise several half-wavelengths, giving modified directional properties. This very fact implies that the system can be used at any harmonic frequency of that for which it is designed, and is thus flexible to some extent.

Practically the feeders are tuned by the insertion of condensers. These compensate for the loading effect of the coupling coil L, and for any error in the physical length of the feeder system, since the 'electrical length' at which this resonates can be reduced at will by variable condensers used in series, or increased if they are used in parallel with the two feeders. By this method feeders of almost any length can be adjusted to resonate electrically as if they were an exact number of

quarter-wavelengths, and the whole system can be adjusted to operate at wavelengths differing somewhat from that of the half-wave radiator *AB*, the node *B* being shifted short distances along the feeder lines by tuning adjustment. The Zeppelin aerial is thus nearly as useful on a number of wavelengths as the end-fed radiator, provided that these are either harmonically related or nearly so, whilst most of the losses through building absorption have been overcome.

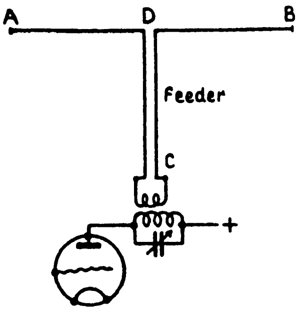


Fig. 158.

For a tuned feeder to be strictly non-radiating it is necessary for it to be strictly symmetrical, both physically and electrically. The two lines are spaced at from 4 to 16 inches apart by suitable insulators, and are supported parallelly and as far from objects as possible. Their capacity to earth cannot be exactly similar in practice, however, and this is particularly the case when the feeder is taken round bends, which must not be of too small radius. Electrically it is difficult to obtain identical current distribution along the whole length of both conductors, particularly near the ends *B* and *D*, for at *D* the current will fall to zero whilst at *B* it will only fall to a low value determined by the impedance of the radiating portion *AB*. Complete cancellation of field is thus not possible, and slight radiation from the feeder will occur. This may represent considerable energy when the transmitter is of high power.

A better balance results if it is possible to feed the aerial at its centre, as in Fig. 158, in which the radiating section *AB* will be one or more half-waves long as before. The two feeder lines are now equally loaded, and a much more nearly identical current distribution is to be expected. Such an aerial can be termed a 'centre-fed half-wave' or a 'centre-fed Hertz' aerial. If the radiating top be one, or an odd number of half-wavelengths long, the feeder will reach it at a current maximum, when the system is termed 'current fed'; whilst if the radiating portion be an even number of half-wavelengths the centre will be a point of potential maximum and low current. The aerial is then termed 'voltage fed'.

As has been stated the tuned feeder line in any form has the advantage that it can be used over a range of wavelengths, but since radiation from the feeder is not quite zero, losses may occur. These are partly due to actual resistance of the conductors and to dielectric loss in the insulators and supports, and largely to residual radiation leading to power absorption by nearby objects. The fact that 'standing waves' occur along the tuned feeder implies that there will be points of high potential and others of high current at which losses will be a maximum. These would be reduced if the feeder current could be constant along its entire length.

Consider either a long aerial wire or a pair of long wires forming a tuned feeder. If energy in the form of a wave be fed into one end of this, it will travel along it with a velocity slightly less than that of light until the far end be reached. The wave will then be reflected back towards the starting end. If the reflection occurs at an open end having infinite impedance, it may be accompanied by very little loss of power, the wave returning to the starting-point with its amplitude only slightly reduced through radiation or resistance. Should the feeder be connected to a load at the far end, then in general a part of the wave energy will be absorbed by the load and the remainder reflected. Reflection can also occur at the end from which the wave originated, and energy may travel backwards and forwards along the wire for some time before it is entirely dissipated.

Clearly a maximum condition will occur when the reflected waves are in phase with and thus reinforce those which preceded them, and when reflection at each end is complete. This condition exists in an open resonant aerial or feeder which is an exact number of half-wavelengths long, and is characteristic of a Hertzian radiator. It results in a very rapid oscillation of the wave energy from end to end of the aerial, but since all these waves are in phase and produce current maxima at the same points along the aerial, the effect is produced of a stationary or 'standing wave'. Permanent current and voltage nodes can be measured at intervals of a half-wavelength, exactly as if the wave had been 'frozen' into position. This is merely a recapitulation in different words of the current distribution within an aerial, viewed from a fresh and important angle.

We have seen that the tuned feeder behaves similarly, being in fact simply an extended aerial arranged to have minimum radiation. Standing waves occur upon it when correctly tuned, and give rise to the losses mentioned. These would not exist unless reflection were taking place at the far or loaded end of the feeder, D. Now a resistance or equivalent load at this point will absorb energy at a definite rate as the wave disturbance reaches it. Similarly, energy will arrive at a definite rate depending upon wavelength and other dimensions. If these two rates be made equal, all energy which reaches the far end of a feeder will be completely absorbed by the load, none remaining for reflection. Standing waves are thus eliminated, and the feeder becomes a true non-resonant feeder line. Current will flow uniformly from source to load, with a slight reduction along the feeder caused by losses, but with no maxima or minima due to resonance. Losses will thus be maintained at the lowest possible value consistent with the amount of power to be transmitted.

Used in this manner the feeder becomes a radio-frequency transmission line, analogous to the lower frequency lines familiar to telephone engineers, or even to the electric supply mains. The rate at which energy passes along the feeder is determined by a quantity known as its characteristic resistance or 'surge impedance', which depends only upon the dimensions of the feeder such particularly as the gauge and spacing of the two conductors. This can be loosely termed the impedance of the feeder. When terminated at the far end by a resistance load equal to this surge impedance, the condition for transmission of power without reflection is found to exist. Fortunately a resonant circuit such as an aerial behaves as a pure resistance to the resonant frequency, and so *at this frequency only* it forms an efficient load into which the feeder can deliver power.

Feeders themselves are constructed according to three main principles. Firstly, there is the pair of parallel conductors, exactly similar to the tuned type of feeder. The surge impedance of this type is given by the expression:

$$Z = 277 \log (b/a),$$

where b is the spacing between the lines and a the radius of the conductors in the same units. Since only the ratio b/a comes into this expression, the actual units are immaterial, Z being in

ohms. This value will only hold when the feeder lines are parallel and perfectly insulated. At bends or points where spacing varies, the surge impedance is also liable to vary, when a degree of reflection occurs. This, and also reduced insulation through damp or exposure, will introduce additional losses into the system above those due to conductor resistance. The latter loss is small, a typical value being 0·5 to 1·0 decibel per 100 feet of feeder at a frequency of 10 megacycles, about 30 metres wavelength.

Secondly, as a modification of this construction it is possible to employ a signal-wire feeder, analogous to an aerial wire, when the second conductor is regarded theoretically as formed from an electrical image of the feeder formed in the earth below it. This system may be convenient and will be described later. Its losses are somewhat higher than the best twin-wire lines, but its convenience makes it valuable for portable or amateur use.

Thirdly, there is the Concentric type of feeder, consisting of an inner conductor of wire or tube surrounded by an outer conductor composed of larger diameter tube. For best efficiency the ratio between the radii of these two tubes should be about 3·6, but for reduced capacity which is desirable at high frequencies the inner conductor is sometimes made of fine wire. This is the case in the concentric cables used for the distribution of television programmes. The surge impedance in this case is given by:

$$Z_0 = 138{\cdot}5 \log R/r,$$

where R is the inner radius of the outer tube, and r the outer radius of the inner tube. It will be seen that the impedance tends to be lower than that of spaced lines occupying a similar total width.

The concentric or coaxial transmission line is favoured in commercial working, since although more costly to erect than the open or spaced wire type, it has a number of advantages. We have seen that twin-wire feeders work best when a symmetrical load such as a centre-fed aerial is connected across them, thus obtaining equal loading and current in each line. This applies to the non-resonant transmission line as well as to the tuned feeder, although there are expedients for minimizing it in the former case. Many practical aerial arrays are not symmetrical, and when this is so the concentric line can be used with the outer tube earthed at a number of points. Even

when the aerial cannot be so treated, it is usually possible to arrange for an earthed outer conductor with the help of a transformer between feeder and aerial.

The inner live conductor is now completely shielded by the earthed outer tube, no external field is likely to exist, and the feeder is strictly non-radiating even when not quite correctly terminated. The inner conductor can be at any high potential without risk of shock to persons touching it, and is protected from the absorption of nearby objects. The whole feeder may even be buried in the ground or run a few feet above it, whilst the 'open' type must be supported clear of objects. In addition the insulators which support the inner conductor are protected from damp or dirt by the outer tube, little deterioration from weather being likely. As a result of these several factors the concentric feeder has the lowest losses of any, and by the use of large diameter copper tubes these can be reduced as far as may be necessary or as cost will permit.

For receiving purposes and light or mobile transmitting work a satisfactory feeder can be made from ordinary twisted flex, or the twin rubber covered power cable used in workshops. This will have a surge impedance of the order of 80 ohms, which happens to be about the same as the impedance at the centre of a half-wave aerial. It can therefore be inserted directly into this point without serious losses by reflection, and although the dielectric losses in the cable will be relatively high, a satisfactory transmission of power results. Special cables having low radio-frequency losses are now being made for this work.

The flex feeder will be substantially fieldless, and, in reception, can be run through intense disturbing fields without picking up serious interference. This is the principle underlying most modern types of noise-reducing or anti-interference doublet or dipole aerials, as they are commercially described. An efficient aerial is fitted up as high as possible, and outside the field of passing traffic, electrical appliances, or mains. These can all set up disturbing fields which will excite an aerial placed near to them, but which is confined mainly to an area near to the ground. One of the several low impedance feeders joins this aerial to the receiving set, and can safely pass near the sources of electrical noise without excessive risk of pick-up.

The increasing use off domestic electrical appliances has led to the need of similar 'anti-noise' aerials for broadcast reception

in urban areas. Since such aerials must receive over a wide range of wavelengths, however, they are in no sense half-wave Hertzian aerials, and have nó definite impedance at the centre. It is usual, therefore, to employ an impedance matching transformer, and to connect this to a point in the aerial having a high impedance at all wavelengths. This condition exists at one end of the aerial, which may be joined to earth through the transformer, working as a Marconi aerial at the longer wavelengths. The transformer steps down to a low impedance suited to the feeder, which may be of screened cable, enclosed in a metal braid covering which is earthed. A similar transformer is used to step up impedance and voltage from the lower end of the feeder to the receiver. Since a correct impedance match is not possible at all wavelengths, the transformer design is a compromise, and there may be a loss of signal strength at some wavelengths. Since, however, the aerial system picks up little noise, the ratio of signals to noise may be improved.

The impedance of the great majority of transmitting feeders will be higher than 80 ohms. The spaced line variety will be found in practice to lie near to the figure of 600 ohms, which is very representative. Concentric types will exhibit a somewhat lower impedance, 400 ohms being a typical value. Whilst these figures will differ and need to be worked out for every practical case, they will lie near enough to the above values to enable us to use these throughout the following paragraphs, in which we must now consider the methods necessary to prevent reflection losses at each end of the transmission line.

We will consider the input end of a feeder first. Here there is actually no question of 'terminating' with a resistance equal to the surge impedance of the line, because energy is being fed in at this point. If it reaches the remote end and is there fully dissipated in a correctly proportioned load or aerial, no energy will be reflected back to the input, and hence no further reflection can occur there. It is thus immaterial from the point of view of the avoidance of standing waves and feeder radiation what form of coupling be employed at the transmitter, the correct operation of the feeder as such being entirely decided by the correctness with which it is terminated at the remote end. In the particular case in which the feeder is not correctly loaded at the output end, then some energy will be reflected back to the input, and re-reflection will be a minimum if the

resistance at that end also equals the surge impedance; but this is merely a precaution to reduce the effects of mismatching at the load. It is thus not strictly correct to say, as is often done, that a non-resonant feeder must be correctly matched at both ends. The output end only is important.

Input matching matters, however, from quite another point of view, that of energy transfer from the transmitter. We have already noted on several occasions the essential condition for maximum transfer of energy from a source to a load—if both are pure resistances it is well known to occur when these resistances are equal. There is then an energy transfer of 50 per cent., half remaining to be dissipated in the internal resistance of the source, or generator, whilst the other half will appear in the equal resistance of the load. This principle applies to the loading of a radio transmitter by an aerial or feeder.

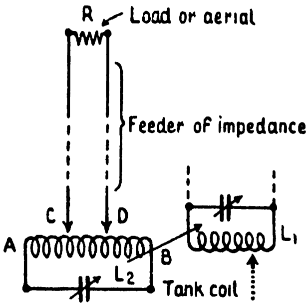


FIG. 159.

Suppose the feeder be tapped directly on to the transmitter tank coil. The surge resistance of the former being R, maximum energy will be fed into it if the impedance of the driving circuit were also R. But feeders seldom exceed 600 ohms, and this is too low a value for insertion into the anode circuit of typical transmitting valves. The tank circuit impedance necessary to match the valve in the manner we have already studied may lie between 2,000 and 10,000 ohms. Let it be Z_t, which will be several times R. Now in the diagram of Fig. 159 it will be seen that the feeder has been tapped across a portion CD of the tank coil AB, thus forming an auto-transformer of ratio CD/AB. We know that the impedance ratio of such a transformer will be the square of the turns ratio, or $(CD/AB)^2$. If this be made equal to the ratio of R/Z_t, then from the point of view of the feeder the tank impedance will have been reduced to a value equal to R. The condition is exactly similar to that studied in the matching of low-frequency output stages to the loudspeaker load, and the condition for optimum transfer of energy is:

$$R : Z_t = CD^2 : AB^2.$$

Fortunately, however, the impedance Z_t is not resistive in this case, but is mainly inductive. The current flowing in it will be nearly out of phase with the voltage, and will be largely 'Wattless', only a small proportion of the product 'volts × amperes' being wasted as heat. Thus only a small percentage of the available power will remain in the tank circuit, far more than 50 per cent. passing into the virtually resistive feeder load. Efficiencies up to 97 per cent. are quite possible at this end of the line.

One advantage of feeder systems over direct aerial coupling which has not yet been stressed is that since the feeder only operates efficiently at the one frequency for which it is designed and terminated, harmonics contained in the transmitter output will be poorly transferred to the aerial. Their suppression is assisted considerably. We have seen the reasons for loose coupling to an aerial, to reduce harmonic radiation. This is equally practicable in feeder working, being illustrated in Fig. 159. The circuit *AB* is inductively coupled to the actual tank coil, or may be coupled by a short sub-feeder, or through any form of harmonic-suppressing filter network, of which several are in use. The outgoing feeder is tapped across the correct proportion of this coupled tuned circuit, as before.

The arrangement has several advantages. For one, the feeder coupling and that to the tank coil can be independently adjusted for the best condition in each case. There is no direct connexion to the tank, and thus high voltages are kept out of the feeder line, and there is no risk of unbalancing a push-pull power stage. Harmonics are still further prevented by the inductive coupling, and it is found in many cases that most of those which remain are transferred by stray capacitive coupling between L_1 and L_2, rather than inductively. If a 'Faraday Shield' composed of an open screen of earthed parallel wires, common at one end but comprising no closed loops to cut the magnetic field, be interposed between the inductances, then this last trace of capacitive harmonic output can be almost entirely eliminated.

To sum up, it may be said that impedance matching at the input end of a feeder is chosen so that power may be transferred efficiently, this being done by any arrangement which enables the feeder to place a correct load upon the tank circuit, whilst the latter maintains efficient matching of the valves used. There are a variety of networks in use which combine these two

functions with harmonic elimination in a single device. We have noted the use of an inductive coupling for these purposes, and it will be instructive to consider one type of filter which does the same things by a capacitive-impedance match. This is shown in Fig. 160 (*a*) for the case of an open aerial or single-wire feeder; if a two-wire feeder is used, the circuit is duplicated

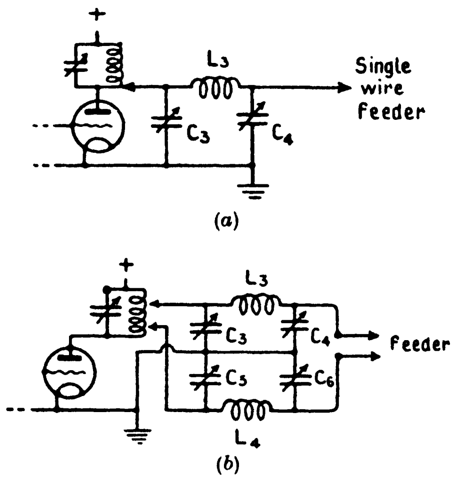


FIG. 160.

as in Fig. 160 (*b*). In conjunction with L_3, the condenser C_3 determines the impedance of the filter as 'looked at' from the tank circuit, and can be so adjusted that the correct load is placed upon it. In use, the tank is accurately tuned to resonance and then C_3 is connected. It is adjusted to give correct loading, the tank tuning being in no way touched to compensate for the alteration that might be expected from alteration to C_3. This is because, in correct adjustment, the filter should behave as a pure resistance load, which will not then alter the tank resonance.

C_4 can now be set to a value which gives maximum feeder current, which will occur when C_4 in conjunction with L_3 forms an impedance equal to that of the feeder. Both sides of the network are now correctly matched, and since the reactance of a condenser varies inversely with capacity, the relationship should be:

$$R : Z_t = C_3 : C_4.$$

If at the same time a suitable value of L_3 has been chosen, the

whole combination of C_3, L_3, C_4 will form a Π-section low-pass filter, having optimum transference to the carrier frequency but very high impedance to harmonics. Physically this can be seen by noting that at increasing frequencies the reactance of L_3 will be increasing, thus impeding the path of current to the aerial, whilst the shunt reactances of C_3 and C_4 will be falling, thus by-passing an increasing proportion of the harmonic energy.

We must now turn to the more important question of termination of feeders at the output end, which for the present will be taken as a half-wave radiator. Harmonic aerials or arrays, being composed of a number of half-wave units, can be treated as the sum of several such cases.

The impedance of a half-wave radiator has been stated to vary uniformly from 73 ohms at the centre up to some 2,400 ohms at each free end. In the simplest case this fact allows us to make a direct connexion to the feeder, since there will always be a point along the aerial at which its impedance equals that of any transmission line. The case of a twisted flex line having a surge impedance between 70 and 100 ohms has already been given. This is about the impedance of the centre of the aerial, which will vary between about those limits according to its distance above ground and surrounding objects. If the aerial be broken at the centre as in Fig. 158 and the feeder inserted, a reasonably good termination results. To reduce the sudden impedance change and minimize reflection if the aerial and feeder impedances are not exactly equal, it is usual to fan out the last few feet of the feeder, thus giving a gradual increase of impedance up to that of the aerial, which is likely to be somewhat greater.

The single-wire feeder forms an ideal example of the next method of termination. It will have a surge impedance of the order of 500 to 600 ohms. A point between 0·3 and 0·4 of a wavelength from either end of the aerial will have this identical impedance. The average value is about 0·36, and may be calculated from the known gauge of wire in the aerial and feeder, or measured. The latter is done after the feeder is attached and in use, the point of attachment being varied until several ammeters placed at points along the feeder indicate uniform current. This shows an absence of standing waves, and is a general method for adjustment of all types of feeder. The single-wire

feeder system was first fully appreciated by an amateur worker named Windom, after whom it is frequently named.

Twin-wire feeders having an impedance near 600 ohms could be inserted into an aerial at the point chosen above, but this is not altogether desirable because the unequal length of aerial on each side of this point would tend to unbalance the feeder.

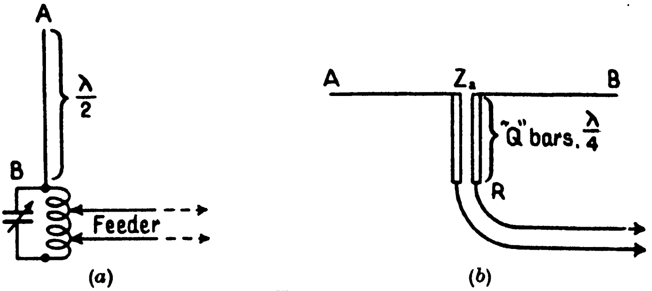


FIG. 161.

A better method makes use of an impedance-matching transformer, exactly equivalent to the tapped tank coil previously mentioned, and by which the feeder impedance can be 'matched' to the aerial at a convenient point, such as the centre or one end of the latter. A concentric feeder having the outer tube earthed demands such a matching transformer, as it is inherently unbalanced.

Fig. 161 (*a*) shows the case in which a transformer matches a concentric feeder to one end of a half-wave radiator, which will represent 2,400 ohms to the feeder. It is clearly a convenient method for the excitation of individual elements of a beam array. Feeders may be forked into branches, so that energy can be distributed to the several units of an array, in which case the change in feeder impedance at each branching point is taken care of either by the insertion of matching transformers, or by a progressive alteration in the spacing of the feeder lines. A 600-ohm feeder must, for example, branch into two each having an impedance of 300 ohms, or if it be made to fork into two identical 600-ohm branches, a 2:1 impedance transformer must be inserted to guard against reflection and power loss.

It is not convenient to hang up coils or matching transformers in the middle of elevated horizontal aerials, and other very efficient types of transformer have been evolved, impervious to

weather. One of the best is composed of two parallel metal tubes, spaced by the order of their diameter by rigid insulators, and a quarter-wavelength long. This is often termed a 'Q' matching section in American publications. These 'Q' bars form a short line of very low losses, so that a certain amount of standing wave energy can be allowed along it without serious

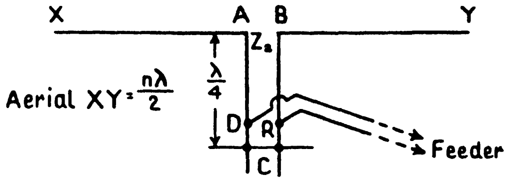


FIG. 162.

attenuation. Its surge impedance can be accurately calculated, and this is chosen to be the geometric mean between that of the feeder and of the aerial at the point to which it is coupled, namely $\sqrt{(R \times Z_a)}$. When this condition exists, and the aerial and feeder are joined to opposite ends of the bars as shown in Fig. 161 (*b*), the loss of power by mismatching at either end is found to be slight, because an increase in impedance at one end of a line above its own impedance is found to be exactly balanced by a corresponding reduction at the other. This is a fundamental condition which holds if the two terminating impedances are chosen from the preceding expression.

Reverting once again to resonant systems, it is possible to match a non-resonant feeder very effectively at one frequency by the use of a short tuned feeder, or 'matching stub', as the terminating transformer. Fig. 162 illustrates this construction, the lines AC and CB being a quarter-wave spaced feeder short-circuited at the current antinode C. This point must be adjusted by a sliding link or the like until resonance occurs, when there will be a uniform change of impedance along AC from that of the aerial at A to almost zero at C. Clearly, therefore, an intermediate point D must exist at which the impedance of the feeder is matched, and if the latter be tapped on at that point, energy will be transferred efficiently into the closed line ACB. Here it will tend to set up standing waves, but these will give rise to little loss since the line is non-radiating, most of the energy passing into the radiating portion of the aerial proper. The action of these matching sections is sometimes difficult to grasp,

but if they be regarded as 'linear transformers' similar to that of Fig. 161 (*a*) but having the turns straightened out into a linear inductance and tuned to ensure the correct current distribution, their function should be more clear. Matching sections or transformers are mainly useful when an aerial is of harmonic length, or an array of several elements. A single half-wave aerial can be more simply end- or centre-fed, as in Figs. 156, 157, and 158.

In conclusion it is interesting to note that a quarter-wave line similar to that just described, joined at one end as at C but open at the other as at A and B, will have a very high impedance indeed to its exact resonant frequency when viewed from AB. This leads to that astonishing device the 'resonant insulator', which will behave as an insulator at one frequency but as a good conductor at most others. An aerial or feeder can be attached at points A and B to a quarter-wave line, the far end of which may be earthed. This forms a convenient method of supporting the aerial mechanically, straining it or steadying a bend against wind. At very high frequencies it is found that the quarter-wave line offers a higher impedance than an insulator of porcelain or glass, the losses of which increase with frequency owing to capacity and dielectric losses. The efficiency of the whole supporting system can thus be raised, with the added advantage for fixed wavelength working that radiation at harmonics or the charges caused by static electricity or lightning can leak away directly to earth. In the same way, a half-wave or quarter-wave open ended line forms a very efficient acceptor or wave-trap circuit, and may be attached to an aerial system to absorb power at some unwanted frequency.

We have dealt somewhat fully with aerial and feeder systems because the aerial is a most fundamental part of any radio station, whether transmitter or receiver. The rapidly increasing use of short waves has rendered the simple Marconi aerial semi-obsolete, and the engineer of the future will be almost exclusively concerned with resonant or directional aerials. Efficient design of a short-wave aerial can make an enormous difference to the performance of a station, and is thus a subject of first importance.

EXAMINATION QUESTIONS

1. Explain how a frame aerial is used for direction-finding. A frame aerial is set up with its plane in the direction of a transmitting station. When tuned, a hot wire milliammeter in the frame circuit indicates 10 milliamperes. If the frame is rotated through an angle of 45 degrees, what will be the reading on the milliammeter?
City and Guilds of London Institute. Preliminary Exam. 1937.

2. Describe, with sketches, giving approximate dimensions, any form of directive or beam aerial and state for what wavelengths the system described is suitable.
C. and G. of L. I. Preliminary Exam. 1930.

3. Give a diagram showing the general distribution of current and E.M.F. in an inverted L receiving aerial. Account for the shape of your diagram. *Institute of Wireless Technology. May 1935.*

4. Explain the principles of the Adcock direction-finder. Why is this system less affected by 'night errors' than a direction-finder employing a frame aerial?

5. Explain the advantages of conveying power to a transmitting aerial through a non-radiating feeder. Describe one form of such feeder, and the precautions necessary in coupling it to the aerial and transmitter, stating why these are required.

6. What do you understand by 'night effects' when obtaining bearings on a loop aerial? Explain their action, and state how they can be eliminated. *A. M. I. W. T. June 1937.*

7. What is meant by the characteristic impedance of a transmission line? Calculate the impedance of a transmission line consisting of two coaxial conductors, the outer radius of the inner conductor being ½ inch, and the inner radius of the outer conductor being 2½ inches. (If this line is terminated by apparatus whose input impedance is $250+j0$ ohms, calculate the loss due to reflection.)
A. M. I. W. T. June 1937.

Note. The last part of this question is not covered by this book.

8. Explain how it is possible to obtain bearings of a distant transmitting station by means of a loop aerial. How can the 180 degree uncertainty in such bearings be eliminated? Illustrate your answer with diagrams. *A. I. W. T. June 1937.*

9. What are the relative advantages of Bellini-Tosi and rotating-loop direction-finders? State the errors to which these systems are liable. *Grad. I. E. E. 1936.*

10. For what purposes was the Robinson direction-finding system evolved ? Compare its properties with the Adcock type, explaining the principles underlying each.

11. Outline the principle methods of navigation by radio, pointing out those suitable for use at sea and those more suitable for use in the air. What do you understand by the 'radio-compass' ?

12. Certain types of short-wave aerial have been given the names (*a*) Zeppelin, (*b*) end-fed Hertz, (*c*) Windom, (*d*) centre- or current-fed Hertz. Sketch an aerial of each type one-half wavelength long. What is the current and voltage distribution in each case ? Mention any advantages you know pertaining to each type.

CHAPTER XV

BASIC PRINCIPLES OF TELEVISION

STARTING from the early methods which employed damped wave trains and permitted Morse signalling at low speeds, we have seen that the adoption of continuous waves of uniform amplitude and shorter wavelengths has resulted in an enormous increase in the range, speed, and reliability of radio-communication. High-speed automatic signalling becomes possible, and the amount of information which can be exchanged between two places in a given time is far greater.

Telegraphic signalling involves only the interruption of a carrier wave, but we have seen also that in so doing it is impossible to prevent the production of sidebands, which extend an appreciable distance into the radio-frequency spectrum on either side of the carrier frequency. Thus high-speed telegraphy tends to resemble a modulated signal, being in fact a special case of this in which modulation is complete, is at relatively low frequency, and has a rectangular wave form.

The application of graduated modulation in place of simple interruption of the continuous carrier wave has led to the development of telephony, by which information can be exchanged between two stations at an even greater rate. We say this not simply because a man speaking can say more words per minute than can an automatic telegraph, since in fact he cannot; but because in addition to mere words the actual tone of voice and the complex wave forms of sound are transmitted. An orchestral performance involves an extraordinarily complex wave form, and may on that basis be said to convey a great deal of information to the listener. Without attempting the difficult task of defining the abstract quantity 'information' or 'intelligence' precisely along these lines, there should be no difficulty in grasping the way in which this term is used.

It will be helpful to analyse the wave forms that compose a telephonic signal into their three component factors, which are frequency, amplitude, and phase. These three completely define any type of radio-modulation at any instant, it having been shown that the most complex wave form can be treated as composed of a number of sinusoidal components, each completely

defined by these three factors. Modulation can thus only convey information having three parameters or dimensions which correspond to these three basic electrical quantities, although several sets of them can be conveyed concurrently.

In the transmission of telephony we convert the frequency or pitch of each tone to a modulating frequency, the loudness of that tone to an electrical amplitude, and the time of its occurence to a phase factor. A number of these tones combined make up any complex sound, no other parameters being possible. It is fortunate that sound seems to be adequately reproduced by a combination of three parameters only, and that phase which is the most difficult to transmit without distortion seems the least important to the human ear.

Telephony has been presented in this manner so that its relation to the more recent development of radio-television can be made clear. Amplitude and phase relationships are restricted to a limited range of from zero to 100 per cent., and we cannot extend their variations indefinitely to accommodate additional information. The only factor that can be extended as far as practical considerations will allow is frequency. It is not far from the truth to state that the amount of information which can be transmitted over a radio-communication channel is proportional to the maximum modulation frequency, namely to the band width. Whilst not a rigid relationship, we can express this fact by stating that the complexity of wave form dealt with is a general measure of the amount of information, and can be represented by the product $(F \times A \times P)$ in which the letters stand for maximum variations in frequency, amplitude, and phase respectively. Of these modulation frequency is the only factor capable of almost unlimited extension. It cannot theoretically exceed the carrier frequency, and in practice should not exceed 10 per cent., so that more information can theoretically be conveyed over a carrier wave of high frequency. The discussion of actual facts concerning modulation in preceding chapters will have borne out this general contention.

The transmission of sight or vision in a similar manner to telephony has long been the dream of many workers. Theoretically it is possible in the form of an extremely complex modulation, but for many years the necessary scientific principles remained unknown or insufficiently developed. Illumination is the essential factor in vision, and any type of scene as viewed

by the eye must be built up of varying light and shade values. Light is thus the factor which corresponds to sound in telephony, the former being a wave motion in the ether whilst the latter is a pressure wave in air. The microphone provides the necessary link whereby sound is transposed into an electrical variation, and a corresponding link for light now exists in the photocell. This device had to be evolved before television became possible, and is one of the factors which prevented earlier development.

It cannot be too clearly stressed that television has not been invented by any one worker, or even by a few leading inventors. It has gradually become possible through the evolution of numerous branches of science to the necessary degree of perfection, radio-communication itself being one of these. Its possibility was more or less clearly seen by several scientists of past generations, and there is little doubt that they would have achieved some form of television had the technical means been known at that time. For the moment, however, we will assume that technical methods are available to perform each step adequately, and survey the problem from its communication angle.

A scene is to be transmitted electrically through the medium of a modulated carrier wave. Of what factors is this scene composed? In the first place there will be variations in illumination or brightness, which can be translated into electrical amplitude by the photocell. Light is also characterized by colour, which corresponds to the frequency of sounds, since the frequency of light waves determines their apparent colour to the eye. Next there is the 'spatial factor' or the relative position of objects within the scene, and this is a variable on account of motion. To these must be added stereoscopic depth, or the impression of three dimensions which we obtain through the use of two eyes in place of a single lens as by the camera.

Let us list these factors which make up a scene. They are:

(*a*) Brightness. (Amplitude of light.)
(*b*) Colour. (Frequency of light.)
(*c*) Spatial Factors.
(*d*) Changing spatial relationships.
(*e*) Stereoscopy.

All five of these must be electrically transmitted without distortion if a perfect impression of vision is to be reproduced.

Unfortunately, however, there are only three parameters available in the normal modulation process, and some form of compromise is essential.

It is interesting to note that even in telephony the same difficulty exists. The width of most communication channels is limited, and we have seen that this makes it impossible to transmit extremely high frequencies. Moreover, the very low tones which approach zero frequency cannot be fully dealt with by most types of amplifying equipment, or by loud speakers. Happily from this point of view the response of the ear to both extremes of frequency declines. Very satisfactory reproduction can be arrived at if the electrical circuits reproduce most of the range which can be heard by average listeners.

In the case of sound amplitude the deficiency is more marked. Sounds ordinarily heard differ in amplitude by over 100 decibels, whilst the maximum variation possible in most communication systems may be some 30 decibels only. An attempt to exceed this will result in either overmodulation of the loudest passages, or loss of the weakest sounds below the unavoidable level of background noise. Some improvement is possible when the latter is exceptionally low, but in any case a scaling-down of volume levels is necessary. It is this fact that imparts such importance to the duties of a broadcast Control engineer. In the case of sound on film recording for example, the greatest care is necessary. Use must be made of the full available volume range without exceeding a sharp maximum limit imposed by the width of the sound track.

A third case is that of binaural hearing. Normally we hear through a combination of two ears of similar sensitivity and by a mental comparison of the sound reaching each arrive at an idea of spaciousness and of direction. This is never catered for by any normal telephonic channel. To do so seems to need the use of two complete channels, each provided with its own microphone and reproducing at one ear only. Very striking experiments have been made along such lines, but there is no reason to expect that binaural hearing will become commercially practicable, or that it is generally necessary. Something must be sacrificed here in the interests of cost and complexity.

The position is very much the same when we turn to stereoscopic vision in either television equipment or the cinema. In the latter case practical methods have been devised, and may

come into general use if their cost is not too great. But the problem of television in three dimensions is considered too difficult, and is ruled out financially. It should be possible by the use of two complete sets of transmitting and receiving apparatus, each dealing with the view seen by one eye; and even simpler methods may exist. At present, however, stereoscopic television can be set aside as a subject for future research.

The list of factors is thus reduced to four by eliminating (*e*) and accepting reproduction in one plane, similar to an ordinary photograph. This simplifies other aspects of the problem, since the 'spatial factor' can now be confined to two dimensions. The position of any point in a plane can be defined by two coordinates, no matter what the system of axes adopted. We will call these X and Y respectively, and they can be regarded as the horizontal and vertical coordinates of any point on our photograph. At any instant, therefore, every point upon the scene can be defined by four parameters only, which are its brightness, colour, X and Y. The changing relationships caused by moving objects will not enter into the problem at a given instant, but they introduce a time factor which must be also borne in mind.

There is still one parameter too many for translation into a single modulated wave. Since the only one not absolutely vital is colour, it is usual at present to dispense with this and to reproduce television in monochrome. It thus becomes strictly similar to the viewing of an ordinary moving film.

There is no theoretical impossibility in coloured television, which has been demonstrated by Baird. The methods proposed take advantage of the combination of primary colours made use of in colour printing or the modern coloured film. As in stereoscopy, two or three complete television images are transmitted by independent apparatus, each representing the light and shade values of the original scene when viewed through a primary colour filter. These are combined at the receiving end to give an impression of fully-coloured reproduction.

Such systems are not easy to design and seldom yield true colour values, owing to the difficulties in obtaining exactly correct primary shades at the receiving end, and in keeping the several primary images accurately matched in both amplitude and position. A more serious drawback, however, is the duplication of equipment and the need for three transmission channels.

Should the three primary images be transmitted over a single channel, it is necessary to increase the sideband width of this three times, or to sacrifice image quality in other respects. This is impracticable at the present stage of development, in which the maximum available channel width will be seen to be barely sufficient for a detailed image of one colour. Ingenious systems have been devised in which colour is transmitted without any very considerable increase in equipment, and over a single existing channel; but it is fundamental in all such cases that a loss of image detail occurs. This is not tolerable at present.

Dispensing therefore with colour we are left with a practicable problem. Only three vision parameters remain, and these can be handled by a modulating system of the usual kind. There are, however, several methods by which this can be done. That generally used is only one of these, chosen for its practicability rather than its perfection. The problem of television will appear most logical if treated as a survey of the several possible basic methods.

We have seen that the illumination at any point within a scene televised can be transmitted as an amplitude of the carrier wave. This seems obvious, and is the only point common to all proposed systems. In general, carrier amplitude will be proportional to illumination at all times, and the minor differences in method by which this is achieved will be mentioned later. Having dispensed with colour and stereoscopy, it remains to transmit the position of every object within the scene in some manner that allows for their changing motions, and through the agency of our remaining modulation factors of frequency and phase. Possible methods of doing this are numerous, but many have been found impracticable.

At this juncture we must introduce the idea of 'definition', using the term in the sense in which it is used in optical work. Since television reproduction will be in one plane and will resemble a plane optical image, it is usual to refer to the received vision as 'the image'. This term is a convenient one whether talking of reception or transmission. In the latter case the term is quite appropriate, because an image of the scene must be formed in a plane by a suitable lens system before it can be electrically transmitted. The definition of an optical instrument is defined by the minimum angle subtended at the eye by two adjacent points within the field of view which can just

be distinguished from each other. This is the angular definition or 'resolving power'. In other cases the actual distance between two such points is stated. In television work other units are more convenient, but there will be no difficulty in appreciating what the term 'definition' conveys in practice. It is a measure of the apparent 'sharpness' of the image, or of the fineness of grain from which it is composed.

Casually considered, vision seems to be a continuous process, but in fact we know that this is not so. The retina of our eyes is a mosaic structure built up from some millions of minute light sensitive cells, each conveying a message of the intensity of illumination falling upon it through a nerve fibre to the brain. A real optical image of the scene at which we look is focused upon this mosaic. Owing to the very large number of cells we do not always realize that there is a limit to the smallness of visible objects, although experience and the effectiveness of the telescope and microscope make it quite clear that the resolving power of the unaided eye is very limited.

In fact the eye can only see as separate objects two points which subtend a distance on the retina not less than that between adjacent cells. This is approximately the deciding factor, and objects smaller or more closely spaced than this will not be distinguishable. By a very wonderful provision of nature a small portion of the retina, the 'area of distinct vision', consists of cells more closely packed than at other parts. The objects we are 'looking at' are focused upon this portion, and are seen with maximum distinctness, whilst those focused upon other portions of the retina are much less highly resolved.

These well-known facts concerning the eye are mentioned to remind the reader that all vision is essentially granular in structure, and is a discontinuous process. That being so, it is improbable, to say the least, that any more perfect process than that used by nature will be discovered, and it is natural to employ a similar discontinuous process in television. A very good idea of the 'definition' or fineness of grain will be obtained from the spacing between cells, or in the event of these being uniformly spaced as in television apparatus, from the total number into which the field of view is divided. We are familiar with the granular structure of printed illustrations, which are made up of from 2 to 5 small 'dots' to the millimetre, varying either in size or number to give the effect of light and shade.

Fig. 163 (*a*) shows an enlarged portion of one such illustration, which gives the impression of continuity if not viewed too closely, and in fact a grain structure of five to the millimetre cannot be distinguished by the average eye without the help of magnification.

It is therefore perfectly reasonable to regard any image as composed of a very large number of very small elemental areas,

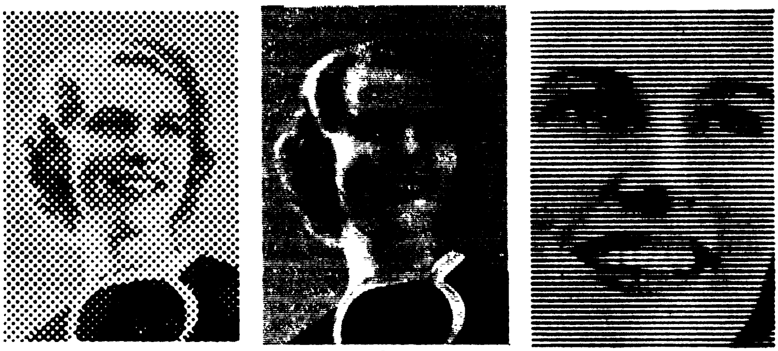

(*a*) (*b*) (*c*)

FIG. 163.

or 'image points', the whole image being regarded as the integration of these. Preferably each elemental area, or 'element', should be so small that the eye is unable to distinguish it individually at the distance from which the image is viewed. In this case there will be no noticeable grain and the scene will appear as detailed and sharp as if actually seen in reality. Since under these conditions the eye can distinguish nothing within the area of a single element, we can treat this as if it were evenly illuminated over its whole surface, and can transmit it electrically as a single value of illumination belonging to that particular element. It cannot be too strongly stressed that this is the identical process believed to take place in the working of our eyes themselves.

In practical television, as in some newspaper illustrations, it may not be possible to employ elements of this smallness. Economy may place a limit to the number of elements used, and if the image is comparatively large, may imply elements which are individually discernible. A very fair reproduction often results even in this case, but we can say that 'definition' suffers, very fine details becoming lost.

The effect of perfect continuity will be regained if the image be viewed from a greater distance, thus reducing the angle subtended by each element at the retina until it reaches the limit of distinct vision. We shall consider presently the actual number of elements necessary to reproduce various types of scene effectively, but we can note now that when this number is less than the ideal there will be an optimum viewing distance, at which all detail existing in the image can be seen but the granular structure remains invisible.

It can be accepted therefore that there is no known television process which is not based upon the idea of dividing the scene or image into elements, each of which can be treated as of uniform average brightness, and the smallness or number of which determines the limiting resolution of which the equipment is capable. This factor, commonly termed the definition of the system, should equal that of the human eye in the ideal case, but falls short of this in practice through technical limitations.

The elements composing a television image may be scattered irregularly as are the cells of the retina, but it will often be more convenient to arrange them in some kind of order. To do this is clearly helpful when one remembers that at the receiving station an image must be built up from elements exactly similar in position to those into which it has been analysed at the transmitter. This can only be done simply if they lie in some geometrical pattern.

A number of arrangements are obviously possible. Many have been tried or advocated, and are capable of more or less satisfactory results. In practice, however, it has been found most satisfactory to arrange them in parallel straight lines, the image being divided up into strips of equal width as shown on an exaggerated scale in Fig. 164. Each strip is regarded as equal in width to one element, and is divided up into equal elements longitudinally, thus forming a mosaic structure of minute squares. This is the simplest method in which we can treat the structure of a television image. It is adequate for preliminary discussion but will need to be extended somewhat later. Fig. 163 (*c*) shows a similar construction, differing only in that the width rather than the density of the strips is varied to produce light and shade. Fig. 163 (*b*) illustrates the excellent definition possible when a larger number of strips can be used.

There is, of course, no obligation to regard the elemental areas as square, and in fact their shape should be of little importance provided that they are too small to be seen individually. In the simpler television systems this may not be the case, and it is found an advantage to employ elements of definite size and shape so that they cover the whole image uniformly, and without overlap or gaps. This gives an impression of continuity, and

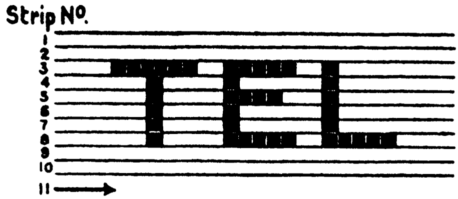


FIG. 164. Image Structure.

makes individual elements difficult to distinguish even when they are large enough to be seen. Square elements are convenient in such cases, but slightly overlapping circles have been used, and accurately placed hexagonal elements are very satisfactory.

The use of elements having width and breadth approximately equal implies that a square image will contain an equal number both along the lines into which they are arranged, or across these. That is to say that the definition will be equal in the horizontal or vertical directions, and approximately equal in all directions. This has been found the most satisfactory compromise when the total number of elements is limited. The use of a rectangular element for example would mean that the smallest detail resolved in one direction would be different from that in others, which whilst it may be desirable in special cases is not the best condition for the reproduction of general scenes from a limited number of elements.

The idea of a mosaic structure gave rise to the earliest systems of television. The most fundamental of all is illustrated in Fig. 165. Here the scene to be transmitted is picked up by the lens L and projected as a real image upon the screen S, which is composed of a mosaic built up of numerous separate photo-sensitive cells. Each of these cells corresponds to an element in our preceding discussion, and produces an electric current or potential proportional to the average illumination

falling upon it. From each cell comes a circuit or transmission channel, which transmits this potential to the corresponding element at the receiving screen. The latter is built up of an exactly similar mosaic, in which each cell is replaced by a light source, the brightness of which varies in proportion to the incoming signal.

Television by such a system will be excellent provided that

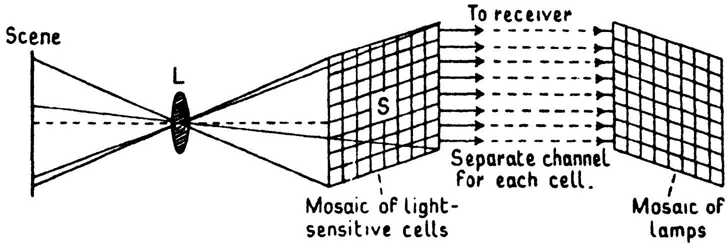


FIG. 165. Early Television System.

the number of cells used is sufficient. About 10,000 would be needed to reproduce average scenes adequately, and since each of these must be connected to the receiver by its own channel, the complexity of the equipment is enormous. Expense renders the method impracticable, but since it is probably the most ideal in all other respects and the key to modern methods, it will repay further study. It was first proposed in about the year 1880 by Ayrton and Perry, but was probably not tested practically so early.

For true reproduction a linear relationship is clearly necessary between the illumination reaching each cell and the illumination produced by the corresponding light source at the receiver. This is equally necessary for the correct reproduction of light and shade in any television system whatever, and implies three requirements. Firstly, the electrical output from the photocells must vary linearly with illumination; secondly, the electrical circuit transmitting this output must be distortionless; and thirdly, a light source is needed which produces illumination proportional to the electrical input.

Many methods have been evolved to meet these requirements, which we shall review as they are encountered. In the above system, however, they are not very difficult to meet. It should also be noted that the principle of balancing one fault against another may be helpful here. If the response of one factor

such as the photo-sensitive cell is not quite linear, this can be compensated for by an equal but opposite departure from linearity in either of the other factors. Thus, for example, the transmission channel can be adjusted to have an amplitude characteristic which will compensate for faults in either the cells or light sources.

The television system we are now discussing introduces no idea of time or speed. This factor only arises when the scene transmitted contains objects in movement. For a stationary scene the illumination reaching each cell will be constant, and so also will be the potential transmitted and the response of the receiving light source. The rate to which these can respond to changes is of no importance, and no idea of band width enters into the design of the connecting circuits. But even when the scene contains movement, the cells need not respond to changes occurring in less than about one tenth of a second, since on account of the defect known as 'persistence of vision' the human eye is not fully sensitive to more rapid changes.

Thus when television came to be tried by this method it was possible to use the Selenium Cell, in which use is made of the changing conductivity of the metal Selenium with illumination. This remarkable effect was discovered accidentally by a telegraphist named May in 1873, whilst employing resistances made of the metal. This cell can be extremely sensitive to small changes of light, but takes an appreciable time to reach an equilibrium position after each change. It is said to possess a marked time lag, and is not suitable for the transmission of very rapid light variations.

The transmission channel between each cell can be a simple wire circuit carrying a direct current of at most slowly varying amplitude, or if a radio wave, it will be modulated by slowly changing potentials giving rise to no extensive sidebands. There is thus nothing difficult to attain here, and it is only the number of channels that complicates the system.

The receiving light source is also not called upon to vary in brightness at any great rate, and it was possible to use ordinary incandescent filament lamps in early experiments. In another case lamps of steady brightness were used, being uncovered by electrically operated shutters. All such devices can be designed to approach linearity without difficulty when working at low speeds, and excellent reproduction is possible. Moreover, there

can be no errors in the correct placing of each element relative to others, or to the corresponding element at the receiver, definition being therefore as great as the number used permits. The received image will be large and bright, excellently suited to public viewing. It is therefore most unfortunate that the large number of channels needed for this television method are not obtainable, for it is the attempts to overcome this fact that have led to most of the difficulties encountered in more recent systems.

Theoretically it should be possible to use this method if the output from each cell can be simultaneously transmitted over a single channel of large sideband width. One method proposed involves a separate oscillator for each cell, generating a low frequency sub-carrier which can be modulated in amplitude by the output from that cell. For N cells there will then be N sub-carriers each of different frequency, and these can be simultaneously used to modulate a single radio wave of very much higher frequency. At the receiver each sub-carrier will be separated by electrical filters, and then applied to control the appropriate light source to which it corresponds. Such a scheme might give excellent results, but although there is no theoretical objection to it, the complexity of the many thousand oscillators, filters, and so on, necessary make it once again useless for commercial purposes.

At present all systems of instantaneous television, namely those in which all elements are transmitted simultaneously over uninterrupted channels, have led to this deadlock. It would be most desirable to employ methods analagous to the human eye, but so far no workable solution has been found, and television would probably have remained an impracticability were it not for that other property of the eye, persistence of vision. It is well known how this defect enables an impression of continuous motion to be built up from the succession of still images projected upon the cinema screen. It also provides a solution to television by allowing the numerous elements to be transmitted *successively* without the eye being able to perceive that fact.

This introduces us to the idea of 'scanning', fundamental to modern television. Scanning simply implies the introduction of the time factor, so that instead of numerous elements being transmitted simultaneously over separate channels, they can

be transmitted successively at different instants of time over a single channel.

Consider a single row of elements in the mosaic screen of Fig. 165. Imagine that the circuits proceeding from each cell are connected to the segments of a commutator, and the rotating arm of this be connected to a single transmission channel. If there are n cells in the row, there will be n segments to the commutator, and if this rotates at uniform speed of M revolutions per second, then the potential from each cell in the row will be successively connected to the channel at M regular intervals each second. At the receiver this process is reversed, the channel being taken to the rotating arm of an identical commutator having n segments each connected to the appropriate light source. If this receiving commutator rotates at exactly the same speed as that at the transmitter, and is strictly synchronous therewith, then at the instant when each cell is sending its particular potential along the channel the corresponding light source will be connected to receive this at the other end. Each circuit is now *successively* in operation just as all were simultaneously before, and provided that the cells and light sources can respond fully during the brief time that each is in circuit, television will proceed unchanged.

The identical process can be carried out simultaneously for each row of cells, rather than for each cell, using a separate channel for each row. Thus in a square image field containing n rows of n elements, the total number of channels is reduced from n^2 (one for each element) to n (one for each row), but the width of each channel will have been increased n times owing to the higher rates of electrical change now being transmitted. This principle has been used in apparatus successfully demonstrated by Karolous in Germany, and has the advantage of retaining a large brilliant screen whilst reducing the number of channels to a possible figure.

We may, however, carry the process still further, applying the circuits from each of the n^2 elements to a giant commutator having this large number of segments, and connecting all successively to a single channel. This was done in the Baird Lamp Screen apparatus by which an image built up of 2,100 elements has been shown to British theatre audiences. Such enormous commutators are inconvenient, and not readily applicable to images built up from larger numbers of elements, but they form

an excellent introduction to present systems in which all elements are 'scanned', and successively transmitted over a single channel by electrical devices.

Scanning therefore means the successive selection of the elements into which we divide a scene, and their successive transmission over a single communication channel. To produce the impression of continuous vision it is now essential to adopt the methods of the cinema, and to transmit every element from 10 to 50 times every second. The elements themselves will then follow each other so rapidly that the eye will be quite unable to detect their motion, whilst the complete images will follow each other at a speed giving the impression of continuous motion, with perhaps a trace of flicker.

Flicker is serious at 10 to 20 scanning traversals per second, (often termed 'frames per second') if reproduction be bright, for the sensitivity of the eye to flicker increases with intensity of illumination. It will be necessary to go up to from 25 to 50 frames per second to overcome this defect completely. Even when that is done the system will be less perfect than the older mosaic screen and 'multiple channel' arrangement first described. There will be possibilities of stroboscopic distortion when objects within a scene move quickly, but it will have the advantage of being workable over a single channel. The enormous rates of change now taking place will imply high modulation frequencies, and the channel will need to be a wide one, using a very high carrier frequency.

The Television Frequency Range. Let us form some idea what these frequencies are. The image field has been assumed square, and contains n lines of n elements, making a total of n^2. If the whole of these have to be transmitted successively M times every second, the total number of elements, each of which must transmit a pulse of potential of its own particular amplitude over the channel each second, will be $n^2 \times M$. Now the opinions of experts differ somewhat as to the time taken to transmit the potential value corresponding to a single element, but since each will produce a pulse similar to one half-cycle of an alternating potential, we cannot possibly assign less than one half-cycle of the highest modulation frequency to any one element.

In other words, since one half-cycle is the smallest unit of modulation that can be treated individually, and one element is the smallest part of our scene that can be so treated, it is

natural to equate these two. On that assumption there will be n^2M half-cycles, or $n^2M/2$ cycles per second, produced as a modulation frequency when the image is scanned and the result transmitted over a single channel. Examination by integration methods indicates that modulation frequencies even higher than this would result in improved definition, but that the best compromise between number of elements and channel width (it is useless to increase one without the other) is given by a maximum frequency lying between the limits of n^2M and $n^2M/2$.

Consider now some actual figures. The first television images broadcast on the Baird system employed 30 lines of elements only, but each of these could be regarded as 70 elements in length, because the image field was rectangular having a ratio of height to width of 7:3. The number of elements was thus $30\times 30\times 7/3 = 2{,}100$. The whole of these were transmitted $12\frac{1}{2}$ times per second, and so the number of elements per second was $2{,}100\times 12\frac{1}{2} = 26{,}250$. If the maximum modulation frequency be taken as half this, or approximately 13,000 cycles, we see that this is rather more than the sideband width obtainable from a typical medium wavelength broadcasting station, which was the transmission channel used.

The light sources and cells used must be able to respond to a change from full illumination to darkness in $\frac{1}{26.250}$ of a second if no definition is to be lost from this cause. Selenium cells and filament lamps would be useless for the work, but fortunately the newer photo-electric cells which we shall describe shortly will respond at this speed. At the receiving end ionized gas discharge lamps, such as the familiar Neon lamp or Mercury vapour tube, can be used. The former gives a fairly linear response up to about twice this speed. In spite of the fact that the system represented about the limit of broadcast modulation at medium wavelengths, the number of 2,100 elements is not a very high one. Actually it proved just enough for good 'head and shoulder' reproduction on small screens not exceeding a foot in height, but was of little value for extended scenes or dramatic productions. In addition the frame frequency of $12\frac{1}{2}$ caused very objectionable flicker.

During recent years steady progress has been made in the use of increased definition, which can be approximately defined by the number of scanning lines into which the image is divided.

By 1930 60-line scanning was in use in America, concurrently with the 30-line system here. Whilst all countries could show experiments at higher definitions, Germany assisted by official Post Office co-operation led the field in public demonstration for several years. At the Berlin Exhibition 90-line images were shown in 1932, and by 1933 had increased to 120 and in some cases 180 lines, the frame frequency being 25 per second. These represent a maximum modulation frequency of from 100,000 to 150,000 cycles for 90-line scanning, and up to 540,000 cycles for 180-line scanning.

Such high definitions will in most cases be transmitted over a low-capacity cable channel, but if handled by radio it will be clear that carrier frequencies in the short-wave region are necessary. Thus for a modulation extending up to 540 kilocycles, or over half a megacycle, the minimum carrier frequency would be about 5 megacycles, or 60 metres wavelength. In practice the fading and generally unreliable propagation of this wavelength would make a considerably shorter one almost essential.

By 1934–5 we find examples of 320 lines in use, whilst in other cases 90 and 180-line images have been increased in frame frequency to 50 per second, thus finally removing flicker from the brighter images. The opinion of German technicians was still almost unanimous that 180 lines at 50 frames was the optimum practicable for radio-transmission, but experiments at higher definitions were going on in most countries in which television had made any headway. About this time the need for a higher definition service in Great Britain was under discussion, and a Committee was appointed by Parliament to select the most practicable standards. Their report was presented in January 1935, and showed that two definition standards had been proposed by leading research organizations, each of which considered its own choice the most suitable. The Baird Company on the one hand advocated 240 lines at 25 frames per second, maintaining that this was if anything slightly beyond the present state of the art; whilst the Marconi-E.M.I. Combine advocated 405 lines at 25 frames, but with an interlaced system of scanning which we shall see is almost equivalent to 50 frames from the point of view of flicker. The image in each case was to be similar in shape to the cinema frame, approximately of the width to height ratio of 5:4; and allowing for this fact the highest modulation frequencies to be contended with would

be in the order of one million cycles for the Baird standard and two million for the E.M.I.

It was decided to establish a broadcasting system which we shall review in more detail later, and which was to employ both systems alternately. Whilst realizing that this would temporarily complicate equipment, it was not felt possible to decide which standard was the most practicable without an extended practical test; whilst the equipment offered by each Company possessed it own peculiar advantages, each likely to be valuable for different classes of work.

The regular radiation of programmes commenced on October 1st, 1936, and was in fact the first regular high definition service in the world, since all previous foreign transmissions had been experimental or irregular. After some six months' experience, however, it was found that the higher definition of 405 lines was providing on the whole better results than the lower, with the additional advantage of much reduced flicker, and it was therefore decided to adopt this standard for all future broadcasting for a perioa of three years.

The explanation of this superiority lies in the fact that in all probability neither standard can be transmitted and received by radio without some slight loss of definition, due to distortion at some point in the long chain of apparatus through which the signals must pass. This distortion may be expected to increase with the band-width needed, and will therefore be greatest in the case of the higher definition standard, but it may well be that the superior image quality possible from this standard will still result in a better received image, even after unavoidable losses have taken their toll. The imperfections of the whole process have of course been considerably reduced even during the brief period during which television has been broadcast, and a strong additional argument in favour of a high definition standard is that scope exists for future technical improvements in equipment. It is not desirable that the standard itself should be frequently increased, since this implies drastic alterations in receivers already in use by the public. The alternative is to employ a standard somewhat in advance of present needs and which will remain adequate for several years of normal progress.

Since the following chapter contains a review of television methods as they stand at the date of publication, we shall not now discuss the broadcasting system further, but will conclude

this chapter with an outline of the development in television methods. So great has been the volume of work done on this subject, and the number of systems developed and for the most part superseded, that it will only be possible to pick out a few of the most representative.

Mechanical Scanning Methods. From the beginning research has centred round methods of scanning a scene as one

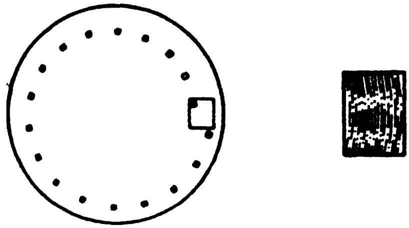


FIG. 166. Scanning Disk.

of the basic problems. We have already seen how a static system has proved to be impracticable, and that a sequential process involving ideas of time and motion has of necessity been substituted. Probably the earliest successful scanning device was the **disk** invented by Paul Nipkow about 1884. It remains one of the most precise, and was until quite recently the most widely used. The scanning disk is simply a rotating opaque shutter, pierced with a spiral row of evenly spaced holes. These are very small, and their size and shape determine those of the elements into which the scene is scanned. The number of holes equals the number of strips into which the scene is divided, the width of the image field upon the surface of the disk being equal to the number of holes multiplied by their diameter. The disk rotates once for every complete scanning of the field, and therefore n times per second for television at a rate of n frames per second.

A portion of such a disk is shown in Fig. 166, and a little consideration will show that at any instant a small element of the field is exposed through one hole only, the remainder being blocked out by the opaque disk. Each hole traverses the field in a line which is in fact the arc of a comparatively large circle, but which becomes sufficiently nearly a straight line when the number of holes and the disk diameter is large. No part of the field is traversed twice in any one revolution, and if the radial spacing between the holes is accurately equal to their diameter,

every part of the field will be explored. The precision of the scanning process is only limited by that with which the disk can be made, and may be very high indeed. The shape of hole used is usually square, but may sometimes be circular, rectangular, or hexagonal as already mentioned. The latter results in the least noticeable demarcation between adjacent strips.

When used for television transmission the disk can be applied in two ways. The most obvious is that in which the image of an illuminated scene is picked up by a convergent lens, and forms a real plane image upon the surface of the disk. Light from that part of the image falling upon the scanning hole at any instant passes through to reach a photo-electric cell, where it gives rise to an electrical potential proportional to the average illumination over that element. This illumination will be rapidly changing as the disk rotates, and the cell must therefore respond with similar rapidity. The selenium cell used in early mosaic experiments is thus unsuitable, and disk television could not be developed until the photo-emissive cell had been evolved to a practical stage by other workers. Changing potentials from the cell are next amplified by valve circuits until they reach sufficient amplitude to modulate a radio-transmitter or to pass safely over wire circuits, and we have noticed that the maximum frequencies produced will be very high when the number of elements is increased. It was thus also necessary to await the development of valve amplification and short wave transmission before effective television became a possibility.

Photo-electric cell is the term used to cover all devices able to produce electrical energy under the influence of light. The transformation of changing light intensities into corresponding electrical potentials or currents is a fundamental need of any television system. We must therefore pause for a moment to consider what types of cell exist, and what are their main properties.

Photo-cells. Photo-cells fall into three main classes, termed 'photo-conductive', 'photo-voltaic', and 'photo-emissive' cells respectively. Photo-conductive cells have the property of changing internal resistance upon exposure to light, the selenium cell which we have already mentioned being the best known example. The most widely used construction is described as a 'condenser type' cell, in which two sets of metal electrodes are used. These are arranged much as are the plates of a condenser,

the small spaces between them being filled with selenium. The element shows little sensitivity to light in its common commercial form, but must be converted into its grey crystalline or 'metallic' condition by a careful heat treatment after it has been deposited between the electrodes. Sensitivity is highest when the selenium is deposited in a thin film of only a few molecules in thickness, and such 'thin film' cells may be enclosed in evacuated glass containers to preserve the film from dirt and oxidation.

The resistance of photo-conductive cells normally falls on exposure to light, the final value being roughly proportional to the intensity of illumination; but unfortunately an appreciable time is taken before the final steady resistance is reached. This 'time lag' limits the response to varying illumination of high frequency, since the selenium will not have time to change fully in resistance during a single 'cycle' of light. The effective sensitivity of the cell may be remarkably high when illumination changes gradually, and is therefore well suited to such work as the control of street lighting or similar commercial uses. The output falls, however, as the rate of change in illumination is raised, and although cells have been designed which will respond fairly well up to several thousand cycles per second, and which can be used for talking-film reproduction, it has not been possible to reach the frequencies necessary for high definition scanned television.

Photo-voltaic cells suffer from similar defects, being unable to respond at high frequencies on account of high internal capacity. Certain of them are very stable and sensitive, being well suited to general commercial uses such as photometric measurements. Types employing a fluid electrolyte are based upon the Becquerel effect. They are termed electrolytic cells, but are little used to-day. The solid electrolyte or electronic types are exemplified by the copper oxide cell, which employs a layer of cuprous oxide upon a copper surface very similar to the copper oxide rectifier. Upon illumination electrons pass from the oxide to the copper to form a photo-electric current.

Photo-emissive cells are the only type widely used in modern television, on account of their nearly instantaneous response. They depend upon the effect first observed by Hertz, that electrons are emitted from certain materials under the influence of ultra-violet light. The modern cell is based upon the work of

Elster and Geitel, who showed that certain metals enclosed in a vacuum can emit electrons under the influence of visible light. The 'photo-electric cathode' seems to resemble a dull-emitting valve cathode in many respects, somewhat similar materials to those able to emit electrons when heated being also photo-sensitive. Those most used are caesium, sodium, and potassium, their oxides, hydrides, and their other compounds.

A typical cell may contain a cathode of caesium deposited by distillation *in vacuo* upon a surface of silver oxide. This will be housed in a glass envelope similar to that of a valve, and containing an anode which may take the form of a wire ring. Electrons are emitted from the cathode in a number proportional to the intensity of light falling upon it, and are attracted to the positively charged anode to form the photo-electric current. The cell is thus very similar to a diode valve, in which a saturation anode current flows, determined by the photo-electric emission from a cold cathode.

Photo-emissive cells fall into two types, in one of which a high vacuum is employed, whilst in the second traces of gas are allowed as in a soft valve or gas-focused cathode-ray tube. The photo-current in a high-vacuum cell is produced solely by emission, and is thus small but strictly proportional to light intensity. Also the emission of electrons is virtually instantaneous, and the cell operates at a speed as great as that of a valve. Its response to high frequencies is limited only by stray capacities, and it is used almost exclusively in high-definition television work. The gas-filled cell generates a much larger anode current for given illumination, since the gas is ionized by the passage of the photo-electrons through it, and a number of secondary electrons are added to these. Since, however, the ionization process takes time, the response at high frequencies falls off for the same reasons as does the beam focusing of a gas-filled cathode-ray tube.

Gas-filled cells may respond well up to several thousand cycles and are used for sound-films and in low-definition television up to about 50-line scanning. As in the case of selenium it is found advantageous to employ a very thin film of the photo-sensitive material, possibly only one molecule in thickness. This construction raises sensitivity and also extends the colour response of the cell, making it more nearly panchromatic.

The response of photo-cells to light of varying colour differs

widely, and must be taken into account in television transmission. Selenium cells respond mainly to red and infra-red light, having little sensitivity in the blue; whilst the copper oxide cell is most sensitive to green light. Photo-emissive cells tend to be most sensitive in the violet and ultra-violet parts of the spectrum, and may be fitted with quartz windows if used for the latter. A sodium cathode responds from the yellow-green through the blue and violet, having a maximum in the ultraviolet. Potassium cells respond with very high senstivity to the blue-violet, but show no response whatever to red light. Caesium in a thick film responds feebly over most of the visible spectrum, with a maximum in the yellow-green region; but when deposited as a thin film upon a copper oxide base it acquires much higher sensitivity over the whole spectrum, with maxima in the red and in the ultra-violet. With a colour-correcting filter to reduce the response in those regions, the thin film caesium cell is well suited to daylight television; whilst its high red and infra-red response make it suitable for use when the scene is scanned by invisible infra-red light. This is sometimes done to protect artists from dazzle, and has been. termed 'noctovision' by Baird. Good infra-red response is also an asset when televising scenes in misty or slightly foggy weather, when it is found that the television camera is less affected than the human eye.

We must now revert to the consideration of television transmission by mechanical methods. We have seen that in one system we must illuminate the scene if this is not already in bright daylight, and form an image of it upon a disk or other scanning device. Light from one element is selected by the latter and reaches a photo-cell of the types just described. This method may be termed direct or floodlit transmission, because the whole scene must be strongly illuminated by floodlamps at *F* (Fig. 167). In studio work the heat and glare are often objectionable and the cost of lighting a serious item.

To overcome this Baird introduced the spotlight system of transmission in which the optical path is reversed. Here the scene is in comparative darkness, although a degree of uniform illumination is permissible since it only produces a weak direct current from the photo-cells, which need not be amplified. Light from an arc lamp is concentrated by a condensing lens *C* upon the back of the disk, intensely illuminating the area covered by the holes, and a portion of this passes through one hole to be

projected as a beam of light upon the scene. It thus illuminates a portion corresponding to one element at any instant, and scattered light reflected from the scene falls upon one or more photo-cells placed in similar positions to the floodlamps of the preceding method. Fig. 167 illustrates the arrangement, and stresses the fact that the optical path is strictly reversed in

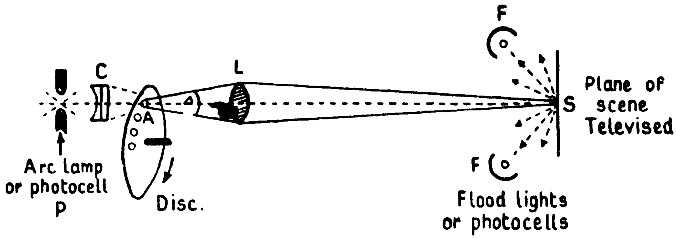


FIG. 167. Simple Mechanical System.

the two cases. These illustrate two fundamental methods of working which can be applied to many scanning systems other than the disk.

In reception a light source is placed behind the disk, and is viewed directly through it. The looker thus sees only the light passing through the single hole at any instant, which corresponds to the element being transmitted. Clearly for this to occur the disks must be identical in proportions, if not in size, and must rotate in strict synchronism at transmitter and receiver. They must not merely rotate at exactly the same speed, but must be maintained in corresponding phase.

The problem of synchronization is present in every form of scanned television, and we shall note methods for obtaining it later. A great advantage of multiple-channel systems would be the absence or reduction of this necessity, but in the adaptation of systems to work over a single channel we have seen that the idea of phase and frequency must be introduced if we are to convey all the essential information.

The receiving light source was in early days very difficult to attain, since those available at the time possessed considerable thermal inertia. For example a lamp filament took time in which to change its temperature and brightness, and thus could not follow the rapid fluctuations necessary for scanned television. This problem was first solved practically by the neon lamp, in which the glow is due to ionization of a rarefied gas

rather than to any form of incandescence. Special lamps were produced having large electrode surfaces over which spread a glow large enough to cover the image field upon a disk. This glow varies nearly linearly with lamp current over most of the working range, a steady 'polarizing' current being provided to maintain a potential between the electrodes permanently above the critical extinction value. Neon lamps can respond with little loss up to some 20,000 cycles per second, and with reduced efficiency up to higher frequencies. The actual response depends very much upon their design and the purity and pressure of the gas filling. Traces of other gases are introduced to improve performance and modify the extreme red colour of a pure neon discharge.

It will be shown later that ample brightness of a received image is very difficult to attain. For this reason the other rare gases which might be used for lamp filling, such as argon or helium in a pure state, were found unsuitable. Whilst their colour may be more pleasant, the intensity of the neon glow was so superior as to make it the standard in most early television equipment. An excellent example of the disk and neon tube receiver was that marketed by the Baird Company in conjunction with the first B.B.C. transmissions.

Discharge tubes of far greater brilliance can be built if the vapours of mercury or sodium be used. These have been employed with moderate success in low-definition television reception, but the heavier gas molecules respond less rapidly to changes of current, and the frequency response of these lamps is inferior with a resulting loss of image detail. Similar remarks apply to the arc lamp, special types of which fed by considerable modulated energy have been used by Baird; and which were used with modulated radio-frequency energy by the revising author in early experiments. Here again the frequency response is limited, and only adequate for low-definition standards.

Optical efficiency. Before considering light sources further we must take note of the major obstacle to mechanically or optically scanned television development, that of illumination efficiency. In the first place it should be realized that in a multiple-channel mosaic system such as those first described, efficiency may be quite high. Light from all parts of the scene transmitted will fall upon its appropriate photo-cell, all available

light collected by the field lens being usefully employed all the time. The only important losses of light will occur through the limited aperture of the lens, and the limited efficiency with which the cells can convert light into electrical energy, neither of which should be very serious. The over-all efficiency of such a system will not be dependent in any way upon the number of cells, since whatever number n of these there may be, each will receive $1/n$th of the available light. Optical efficiency is therefore basically independent of definition, except for minor practical factors. This argument is important, as it is on similar lines that the increased efficiency of modern electronic equipment has been obtained.

The moment scanning is introduced, however, a very serious loss of illumination occurs. At any instant only the light corresponding to one element of a scene is being utilized. In a disk receiver, for example, light reaches the eye through one hole only, that falling over the rest of the image field being blocked by the opaque disk. No matter how scanning is achieved this loss is inherent in the principle. Thus if there are n elements into which the image is successively scanned, the optical efficiency will be reduced by the factor $1/n$, and only this fraction of the total illumination can be utilized at any instant to provide an electrical signal. In transmission the result of this loss means that the scene must be intensely illuminated in the first place, unless the somewhat more convenient spotlight method is used, when a very intense scanning beam produced by a powerful arc light can be used.

Even these steps do not produce good illumination at the photo-cells when n becomes very large. The photo-cell output tends to fall below the limit at which valve amplification is possible, the vision potentials falling to the level of valve and circuit noise. This limit occurs at a definition somewhere about 500,000 elements, or 100-line scanning at 50 frames per second in the case of studio television, although it depends very much on skilful design of equipment. It can be safely said that high definitions of several hundred lines are impossible by the simpler forms of mechanical scanning such as the disk. An exception to this may occur in the transmission of films. A greater intensity of light can be used without damage to the moving film than would be possible in a studio, and higher definition is therefore possible.

In reception low optical efficiency results in an image of poor brilliance, and entirely prevents the projection of this on to a large screen. The goal of television on the cinema screen is still largely unsolved except at low definitions which result in too crude an image for entertainment purposes. However much equipment is improved there seems no possibility of it except by methods which incorporate some of the advantages of the mosaic system. We shall note later that progress is being made in that direction. For direct vision of the image, such for example as through a disk, the limitations are not so serious; but both brilliance and the necessity for a reasonably small scanning device demand a comparatively small image.

Accepting for the moment the losses inherent in scanning, which for a square image field equally resolved in both dimensions can be taken as inversely proportional to the number of scanning lines N^2, we must also take account of optical losses due to the actual scanning equipment. In the attempt to reduce these a variety of methods have been evolved. We cannot analyse them in detail in a work of this scope, particularly as most are now superseded, but will review them briefly as stages in the development of the art.

The disk although simple and precise, becomes very inefficient at large numbers of lines. Firstly, there are mechanical difficulties. Since the number of holes must be increased, either a larger disk or a reduced size and spacing of holes is necessary. The former soon reaches practical limits, since a disk rotating at 1,000 or more r.p.m. cannot exceed a few feet in diameter without great inconvenience. If on the other hand the diameter of the holes is reduced, it becomes difficult to construct them accurately, and the quantity of light which can pass through falls rapidly. In the limit when the hole diameter becomes comparable to the wavelengths of light, diffraction and scattering set in. These holes are easily filled with dust, and if they begin to resemble 'tunnels' through a relatively thick disk, their 'optical aperture' becomes less than their actual diameter. No more need be said to show that the disk has limitations which make it unsatisfactory above some 100 lines at most. Mathematical analysis of a particular set of working conditions and dimensions shows that if other factors can be treated as constant, the optical efficiency of a disk falls off as the fifth power of the number of scanning lines; and this figure indicates at once a very definite

upper limit beyond which the intensity of light passing through will be negligible.

Thus poor illumination prevents the use of the disk at high definition. In one case, however, it remains effective, although it is strictly no longer the same device. For film transmission a disk pierced with a single circle of holes in place of the original spiral may be used. The effect of the radial spacing between successive lines is obtained by a uniform motion of the film itself in the radial direction. Thus the disk traces out a succession of scanning lines across the film, whilst the latter moves longitudinally by the width of one line between each traversal. Since the film must be moved forward in any case, it is simply a matter of substituting uniform for the usual intermittent movement of the film, as is done in sound track reproduction. The advantage of this method lies in the fact that the disk can now contain any number of holes, provided that its speed is correctly related to that of the film. For example, a 240-line scan is obtained from a 60-hole disk rotating at four times its normal television speed, thus traversing the film 240 times whilst a normal 240-line disk would have rotated once. The relationship can be seen to be:

$$N \times F = A \times R,$$

where

N = number of lines into which image is scanned.

F = number of frames of film scanned per second.

A = number of holes in the disk.

R = rate of rotation in revolutions per second.

The left-hand side of this relationship represents the number of lines scanned *per second*, and the right the number of holes which sweep across the film per second, which must of course be equal.

It is now possible to use a smaller disk pierced with a practicable number of holes of adequate size, but it is necessary to increase the speed of rotation. This is possible in practice, and certain recent German equipment employs disks running in evacuated enclosures at speeds between 6,000 and 12,000 r.p.m. The quantity of light reaching the photo-cell has been increased, partly through the use of larger holes, and partly because of the intense illumination possible in a film projector; and it is found

possible to transmit excellent images up to at least 240-line definition. Since, however, only film can be handled in this way, it is necessary to photograph the whole of a vision programme on to film as a preliminary to television transmission.

'Intermediate film' television has received considerable attention, particularly in Germany. At one period of development the only satisfactory method known for high-definition outside broadcasting was to film the event, process the film very quickly, and transmit it whilst still wet by disk equipment of the type described above. Highly evolved equipment by Telefunken was housed in a van, the film camera being used on the roof, whilst the whole process of development was reduced to the astounding time of 30 seconds. A technical difficulty exists, however, in delaying the accompanying sound programme by this time interval, and it is necessary to record the sound also, with a probable loss of quality. The whole system represents a triumph of ingenuity, but is clearly very inconvenient in comparison with the more recent electronic television cameras which operate instantaneously.

As a method of reception the intermediate film provided until recently the only known means of filling a full-sized cinema screen at high definition. It is possible to provide a very small bright image either by disk or cathode-ray methods, and this is suitable for exposure on to film. The film is rapidly developed, and passed through a standard film projector, after which it may be stored for record purposes, or re-emulsified for future use. Here again the accompanying sound must also be delayed.

For direct studio or outdoor work, it was necessary to find scanning systems in which a larger proportion of light can be utilized. One line of development resulted in the lensed disk, in which the small holes are replaced by a number of identical convergent lenses. Prisms are used in another type of construction. At first sight it might be thought that the larger area of these lenses would enable considerably more light to pass, but practically this is not the case. In order to form a sharp and correctly shaped scanning element of sufficiently small area, a 'stop', or opaque plate containing a small aperture must be included in the optical system. Such a stop limits the illumination in exactly the same way as did the small holes of the original Nipkow disk.

To explain fully the performance of scanning systems we should go into optical considerations outside the scope of an elementary electrical text-book. The essential factors of the problem are indicated in Fig. 167. The quantity of light from a scene or image S reaching a photo-cell P (or inversely in the spot-light type of system) will be determined by the *whole* optical path which the rays must traverse. This must always include certain essential parts, which are a 'stop' or small aperture A, and one or more lenses L which form an image of that aperture in the plane of S. It is this image of the aperture A which actually constitutes the scanning element. There may also be other lenses, mirrors, or devices to deflect the rays and build up the scanning motion, but these are not essential to the argument and are not shown.

Now the lens L is not infinite in size, and hence cannot collect all the light from A. It will in fact collect a proportion of that light determined by the solid angle Δ subtended by the lens at A, or a proportion $\Delta/2\pi$. Students of photography will recognize this factor as related to the photographic aperture of the lens.

Secondly, there will be a limitation due to the small stop of area A. This can be most easily understood by considering a case in which the rays of light pass in the opposite direction, such as a spotlight scanning transmitter in which P would be an arc light and S be the scene scanned. Since the path of light is always reversible through a lens system, it is immaterial which case be taken, the total losses being identical. If the intensity of illumination of the stop be now I candles per sq. cm., and its area A sq. cm., then the quantity of light passing through it will be IA units. This is limited by the actual size of the stop, which in its turn is fixed by the size or scanning element we wish to use, and must necessarily be very small for high-definition working.

Thus there are two optical factors which limit the illumination obtained at P from S (or vice versa). These are the limited size (optical aperture) of the lens L, and of the stop A. There may be other losses, such as those due to imperfect reflection at mirrors or imperfect transparency of lenses, but these will be comparatively unimportant. We can now see why there is little difference in efficiency between the perforated or lensed disk. In the former a fixed lens is employed whilst the stop becomes the small hole in the moving disk, whilst in the latter

the stop is fixed and the lenses move. The optical system is essentially the same, and it will be solely a matter of practical dimensions that may make one or the other superior in any particular case. At high definitions, however, the lens disk becomes definitely inferior, because it is impossible to accommodate a large number of large lenses around the moving disk. Small light lenses must be used, and the optical aperture of these will be small, making the efficiency lower than for the Nipkow disk used with a larger fixed lens system.

The Mirror Drum. Several scanning systems have been evolved in which moving mirrors effect the exploration. The optical system is very similar to that just described. There must be a stop as before to determine the size and shape of the scanning element. This is fixed and its image is projected on to the image plane S by a fixed lens. At any convenient point along the optical path are inserted mirrors, which sweep the projected image of the stop over the plane S, thus scanning it as before. In essence we have merely replaced a moving stop or lens by a moving beam of light, deflected by moving mirrors.

The simplest arrangement developed by Mihaly and others employed two mirrors vibrating at right angles. One of these vibrates at the 'line frequency' and sweeps the beam in lines across the image frame S; whilst the other vibrating at the 'frame frequency' simultaneously displaces each line by its own width relative to the next, thus spreading the lines into a 'raster' over the area. Whilst attractively simple, the arrangement has defects which have never been fully overcome. Naturally the mirrors would vibrate with simple harmonic motion, and if this is allowed, the rate of scanning is not constant but sinusoidal. This leads to unequal illumination and definition at the edges of the field relative to the central portion, and is unsatisfactory. Attempts to vibrate mirrors with constant velocity involve such enormous accelerations at the ends of each swing that they have proved mechanically impossible except at low speeds. In either case it has proved most difficult to synchronize the transmitting and receiving mirrors, and to maintain them identical in frequency, phase, and amplitude.

Rotating devices such as the disk are more easily synchronized, since it is only necessary to maintain their angular velocity identical by the use of suitable synchronous motors run from alternating current mains of definite frequency. The phase,

once correct, will remain so. Hence a similar arrangement making use of mirrors has been widely adopted and has given good results. This is the Mirror Drum, in which a number of mirrors equal to the number of scanning lines are arranged round the periphery of a rotating drum or wheel. Each of these mirrors is inclined at a slightly different angle to the axis of rotation, whilst in the other plane its surface is normal to a radius. As a result a beam of light falling upon one point of the circumference is projected by each mirror in turn to trace out a scanning line, and the differing angle between them results in the lines being successively displaced to form a raster.

The mirror drum, invented by Weiller in 1889, has proved one of the most successful of the simpler mechanical television methods. Its method of use is shown in Fig. 168. Light from a source illuminates a stop A, the image of which is projected by the lens L upon the image plane S exactly as in previous methods, but after reflection from the mirrors of the drum at M. The optical conditions are thus basically the same. It is found necessary if the field is to be evenly illuminated to provide a cone of rays sufficiently large to cover two or three mirrors of the drum at any instant. This ensures that the active mirror is always completely filled with light, but it also implies a wastage of that light which falls on the two adjacent mirrors. This is a loss unavoidable in simple mirror drum equipment.

We have stated that the light collected by a lens, and hence the efficiency of an optical scanning arrangement, is proportional to the solid angle Δ subtended by that lens at the light source. This is a fundamental fact, and since it is the principle variable factor involved, it forms a measure of the relative efficiency of scanning systems involving a lens. It so happens that for practical dimensions of the various parts, mirror drum working will often allow a larger cone of rays than the disk systems. Also since the mirrors merely deflect the rays, as large an image as is desired can be built up at large distances from the drum, and the image size is not so closely related to that of the scanning mechanism as it was for the disk. As a result of these features the drum may exhibit a better actual efficiency in use. It was employed in the Baird 30-line receivers for example, where it provided an image as large as $14'' \times 6''$. Above about 100 lines, however, even the drum becomes unwieldy and provides very poor illumination; mainly because the decreasing angle between

successive mirrors as their number is increased results in a small cone angle. Also the distance from mirror drum to image becomes excessive, and cannot be contained in a normal cabinet.

A material improvement is possible by the use of multiple reflection, introduced for this purpose by R. Wilson and the revising author. Here the beam is reflected twice or more by each

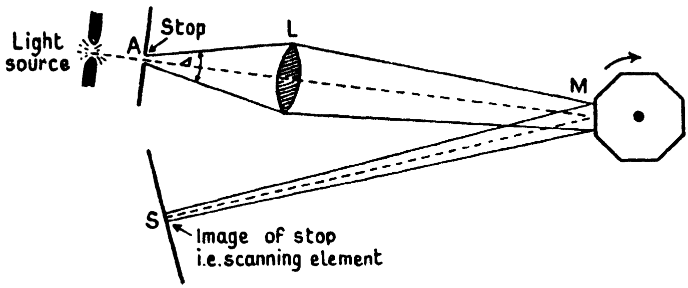


FIG. 168. Mirror Drum Scanning.

mirror, and the angle of deflexion consequently multiplied by two or three. This increases the solid cone angle by four or nine times respectively and increases the illumination proportionally. Alternatively the dimensions of the equipment for a given illumination can be reduced. This principle can be applied to other systems of scanning, and is one of those by which development towards higher definition is possible.

An interesting modification of the mirror drum is that due to Mihaly, in which the ring of mirrors is fixed and the beam is swept over them by a single rotating mirror. The mirrors are arranged round the inner circumference of a hollow drum, at the centre of which is the rotating mirror. An advantage of the arrangement lies in the use of a small rotating part of low inertia, whilst the heavier mirror ring remains stationary. It can therefore contain a larger number of mirrors, each of increased area, and the adjustment of these is simplified. A ray of light passing through this system strikes the mirror ring first, then passes to the central mirror from which it is again reflected back to the ring. An increased cone angle is thus attained through an application of double reflection.

The Mihaly system was of German origin, and has been extended and improved in this country by Traub. Taking advantage of a line-multiplying effect, the complete mirror ring

has been replaced by a part circle containing as few as five mirrors. Since each of these must be correctly adjusted in both planes to make an accurate small angle with its neighbours, it is a great convenience to keep their number small. The central mirror is now replaced by a regular reflecting polygon of perhaps nine faces, not unlike a mirror drum. This rotates at considerably increased speed so that each face operates several times for one scanning of the image. The necessary larger number of lines is thus built up by multiplication of three quantities, namely the number of mirrors in the part-ring, the number of faces on the central polygon, and the number of rotations which the latter makes per raster. Combined with the advantages of an improved light valve, the Mihaly-Traub system has given excellent high-definition results.

We cannot devote more space to the simple scanning systems in which a single moving part performs the whole process. A large number exist of which those mentioned are typical, but all share the defect that at high definitions their optical efficiency falls below the practical minimum. At numbers of lines below about 100 they may be useful for restricted purposes.

The trend of development. A step forward was made when it was realized that the use of two separate scanning devices of the types we have described, but each operating in one plane only, results in much improved illumination. Of the many combinations, we will select 'crossed mirror drums' as the most typical. Here two drums rotating in planes mutually perpendicular receive the beam in turn. Each drum, however, is fitted with mirrors parallel to the axis instead of progressively displaced, there being an angular displacement only in the plane of rotation. One drum will be fitted with mirrors equal to the number of scanning lines, or a sub-multiple of this if the speed of rotation be increased accordingly. It will sweep the image with the necessary number of lines; but each would fall exactly upon the others were it not for the second drum. The latter carries only a small number of mirrors and rotates at low speed, so that the number of mirrors acting on the beam per second equals the frame frequency only. Each mirror thus deflects the beam perpendicularly to the scanning lines, and continues to do so for the duration of one complete frame, so as to place them parallel to each other. We now have a combination of two scanners, termed the 'line' and 'frame' scanners respectively.

The former is concerned only with the tracing of the lines, and the latter only with their displacement into frames.

As a result of this separation of work, each drum can be made optically more efficient. The width of the mirrors parallel to the axis is not now connected with the scanning process, and so can be increased to accommodate more light. In the case of the line-scanning drum we can reduce the number of mirrors, thus allowing them to be made larger, whilst increasing the speed of rotation to bring up the total number of lines to its correct figure. The frame-scanning drum rotates perhaps once per second only, and so may be large and of considerable inertia. Moreover, each drum is only limited in cone angle in one plane, and can be enlarged in the other. The final result is a greatly increased cone angle for each drum, and a larger 'optical aperture' for the whole system, together with many practical simplifications. Development along these lines has made mechanical scanning possible up to at least 240 line definition, and research is still in progress.

The elimination of aperture limitations in one of the two dimensions concerned underlies the effectiveness of two very interesting systems which must be briefly mentioned, those of Scophany and TeKaDe. The former employed an 'echelon prism', whereby the image field was automatically divided into strips by a static optical device involving a stepped construction; whilst the second employs a series of mirror strips arranged in screw-like formation. This most ingenious device has been called the mirror screw, and has given extremely fine results at medium definitions. At high definition, however, it becomes unwieldy. It has been claimed that each is fundamentally the same, coming under the Scophony inventions of Walton.

In operation the essential point is that at any instant a whole scanning line can be regarded as illuminated, the moving element serving merely to select successive points along this line. For example, in the disk it is necessary to illuminate the whole image area, thus wasting much light. If only a single line need be so illuminated the proportion of useful to wasted light is reduced by a factor n, equal to the number of lines. It can be shown that in systems of this kind the optical efficiency falls off only as the third power of n. Here again research is in progress and high definition seems at least a possibility.

Probably the most important step in improving optical efficiency by increasing the cone angle is that evolved by Scophony under the name of 'split focus'. This has been foreshadowed in preceding paragraphs, when we noticed that a mirror drum used to scan in one plane only can employ mirrors extended in the other dimension. Their area is thus increased and a larger cone of rays can be dealt with.

The split-focus principle treats the optical paths in the two scanning planes as independent. The lenses used are plano-cylindrical, having magnification in one plane but no power in the other. The cone of rays is first brought to a focus in one plane only upon the first scanning drum, in the form of a narrow strip of light. In the older systems it was not possible to focus the beam actually upon the mirror surfaces, because of the fact that a mirror produces no deflexion in a beam focused upon its surface. The scanning device must therefore be placed at some other point in the beam, at which the rays are in the form of a larger cone. If, however, the mirror deflects the beam in one plane only, there is no objection to focusing upon its surface in the other plane; and Scophony does this firstly in one plane at the first drum and subsequently in the other plane at the second drum. The beam is thus focused independently in the two planes, focus being 'split' or divided so that the beam can be focused upon each drum in turn in the plane that will not prevent correct operation of that drum.

The dimensions of the drum now cease to limit the transmission of light in the plane in which the beam is focused, for in that plane the beam now has very little width. Neglecting aberration it would be focused into a line of light, of infinitely small width, and could be reflected from a mirror of negligible area. Thus the use of split focus enables each drum to be reduced considerably in size, or alternatively it allows a larger quantity of light to be handled by a given drum. Theoretically the maximum gain possible by this modification is given by a factor n, equal to the number of lines in use.

Light Valves. We have reviewed mechanical scanning, noting some of the difficulties that have been or may eventually be overcome. Not the least of these, however, has been that of a suitable receiving light source, capable of responding to the immense frequency range of high definition television signals whilst providing ample illumination. Primary sources such as

the neon lamp have proved incapable of further development, and it has become essential to fall back upon forms of light valve, able to control the intensity of the beam from an arc lamp or other steady intense source.

Most important of these has been the Kerr cell. In this device the beam of light is polarized by a Nicol prism N, and then

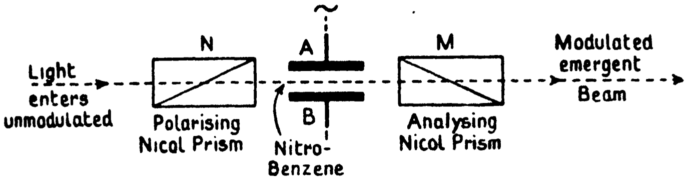


FIG. 169. Kerr Cell Light Valve.

passes through an active material such as nitro-benzene, which forms the dielectric between two plates of a condenser across which the signal potentials are applied. Under electric stress nitro-benzene has the property of unequally retarding the ordinary and extraordinary rays, and of producing an effect very similar to a rotation of the plane of polarization. The emergent light is now analysed (again passed through a polarizer such as a Nicol prism) and the quantity remaining will be dependent upon the potential across the electrodes A and B. If for example, the polarizer and analyser are 'crossed', no light will pass through them. The insertion of nitro-benzene under electric stress destroys this condition, and a quantity of light nearly proportional to the applied electric stress will pass through. The essential parts of such a light valve are shown in Fig. 169.

Unfortunately the optical efficiency of the Kerr cell is also poor, much light being lost in transit through it. The reader may find it interesting to consult other works for fuller details of this interesting device, which possesses many modifications. Improvements have been made in several respects, such as by the use of double-image prisms in place of Nicols, and in the utilization of both the ordinary and extraordinary rays.

A very striking development has been made recently by Scophony in the evolution of a new form of supersonic light valve, in which the Kerr cell is replaced by a bath of oil in which is immersed a quartz resonator. This is excited by radio-frequency energy modulated by the television signals, and

throws the oil bath into supersonic vibration at the frequency of the crystal. Waves of compression pass across the bath, modifying the optical properties of the oil correspondingly, so that it takes on the properties of a diffraction grating. A beam of light passes through the bath, and is brought to a focus in the form of a thin line by a cylindrical lens. At this point is inserted an opaque bar, which completely cuts off the light. In the absence of vision signals, therefore, no illumination passes through the device.

As supersonic waves pass through the liquid, however, diffraction effects occur and destroy the sharp focus of the light beam. This now spreads, and some of it will pass round the opaque bar, to be collected into a new modulated beam by a collimating lens. The amount of diffraction and consequently the light which passes varies with the modulated energy applied to the crystal. This energy need not be large, and response is very rapid since the crystal frequency is chosen to be much higher than the vision modulation.

The beauty of this system lies, however, in the fact that it is found possible to synchronize the rate at which these waves pass through the oil with that at which the scanning beam also traverses the oil bath. Thus it can be arranged that the beam always passes through that portion of the oil at which exists the wave front produced by the vision modulation corresponding to that particular scanning element; and each element can be illuminated for a longer time than usual. We may say that this light valve remembers the illumination of each element for a time, and continues to illuminate it for perhaps the duration of some 500 elements.

It is thus possible to illuminate 500 elements simultaneously, each with its own correct value of light, because the beam pertaining to each element follows the compression waves across the oil bath at exactly the correct rate to remain in its correct position relative thereto. An effect of this complexity and ingenuity is not easily explained in a few words without going into constructional details, but it serves to indicate another line of research which is leading to high-definition reception by mechanical methods. It is probably the first successful attempt to obtain the effect of 'light storage', without which mechanical scanning is unlikely to compete with the newer cathode-ray methods.

It will perhaps be thought that too great stress has been laid upon the defects of the systems outlined, and that television by them would indeed be an unsatisfactory procedure. This is in fact largely the case. The mechanical-optical systems have played a very large part indeed in the development of television, and have yielded good results at limited definitions. Some understanding of them is very necessary and they form the most obvious line of approach to the whole problem. In general, however, they have so far proved unsatisfactory, but it should not be overlooked that research is still in progress. Satisfactory high-definition results have undoubtedly been shown during recent months, and in the field of large screen projection at least the mechanical systems are thought by many to have an important future.

A high-definition television service for home use has only become possible through the comparatively recent development of electronic television, and at the moment this is synonymous with modern methods. The major problems of mechanical scanning have been poor optical efficiency, inertia of the moving parts at high speeds, synchronization and lack of efficient light sources capable of response to very high frequencies. All these defects disappear in the cathode-ray tube, upon which present types of television receivers are based. Here the moving part is an almost inertialess beam of electrons, the intensity of which controls illumination and is in turn easily controlled electrically. The optical efficiency can be quite high, because the time factor enters quite differently into the problem than it does when actual light rays are considered. The electron stream can excite intense fluorescence in a screen even if it acts for an extremely brief period, and the resulting after-glow may last for a prolonged period, resulting in an almost permanently illuminated image however fast the scanning beam may be moving. In only one respect does there seem to be an obstacle. It has not so far proved possible to construct cathode-ray tubes comparable in size to the cinema screen, or to project images from them up to any such size.

Flicker and Interlaced Scanning. The problem of flicker is met with in both mechanical and electronic television systems. We have seen that in the former case it is necessary to increase the frame frequency to about 50, a larger number being necessary as the brightness of the received image is raised.

This is due to the increased sensitivity of the eye to flicker as illumination is increased. Since a frame frequency of less than 25 is ample to give the impression of smooth motion, a large increase in scanning speed and sideband width is needed merely to overcome flicker. In the cinema only 25 film frames are projected every second, but the illumination is interrupted three times during each frame by a three-bladed shutter. Thus the screen is alternately illuminated and darkened 75 times each second. The effective flicker frequency is 75, which is invisible to the eye even at the highest illuminations.

A similar economy would be most helpful in television reception also, but it cannot be achieved by the same method. A television screen is never illuminated all over at one time, but by a travelling spot of light. It is therefore impossible to interrupt the illumination as a whole, since the image has no complete existence at any particular instant. A similar effect can be reached, however, by the method termed 'interlacing' or interlaced scanning. We have mentioned that this is used in the present London broadcast service. Its purpose is to reduce flicker by raising the effective frame frequency, but without any corresponding increase in the actual number of scanning traversals which take place.

To effect interlacing the scanning element traverses half the total number of lines first, omitting every alternate line. Thus the first, third, fifth, &c., line would be scanned, until the whole surface of the image had been covered in a network of spaced lines. The process is then repeated, scanning the remaining second, fourth, sixth &c., lines; which of course interlace with the former to complete a full raster. In this way the screen is covered twice during the period of one complete raster, and if as in the London service 25 frames are scanned per second, the screen as a whole will have been evenly illuminated 50 times.

Interlacing is not difficult by mechanical methods. Thus a disk may be pierced with two spirals of holes, placed so that the two sets of lines interlace with mathematical precision. Triple interlacing is equally possible by this means, and was successfully used by Sanabria in America during quite early years. A 45-line image was scanned by three 15-hole spirals, with much reduced flicker and improved detail. The 405-line London transmission is scanned in two operations of 202½ lines each. The use of a half line is convenient when receiving on the cathode-

ray tube, for it has the effect of automatically displacing each traversal by the width of one line. In this case interlacing is set up in the transmitter, and the receiving beam forced to follow a similar course by the automatic synchronizing circuits, which control each line.

Electronic television differs so much from the mechanical systems, although following the same basic principles, that it will form the material for a final chapter. We shall not attempt to trace the development of this subject, but will combine an explanation of its use to-day with a description of the London television service as broadcast in 1938. This will provide a good example of television by the most advanced forms of equipment.

EXAMINATION QUESTIONS

1. Describe the construction and action of (*a*) a gas-filled, (*b*) a vacuum photo-electric cell. Compare the advantages and disadvantages of each for television purposes. What conditions govern the choice of a cell for (1) spot-light scanning, (2) floodlight scanning?
Institute of Wireless Technology. May 1935.

2. Compare the advantages and disadvantages, from as many aspects as possible, of a mirror drum and a Nipkow disk, (*a*) for 30-line scanning, and (*b*) for 150-line scanning, when these are employed for television transmission. *I. W. T. June 1934.*

3. Why has it been found that the use of wavelengths of the order of 5 metres is necessary for the successful radio-transmission and reproduction of television images having, say, 200 lines, 1 : 1·5 ratio, and image-repetition frequency of 30? What is the maximum modulation frequency of such a system? *I. W. T. June 1934.*

4. Outline the reasons why mechanical scanning becomes unsatisfactory at high television definitions. Mention lines of development by which this difficulty may be reduced or overcome.

5. Explain the differences in design and properties of an amplifier suitable to television, as compared with one suited to telephony only.

6. Describe the intermediate film system of television transmission. What are its advantages and disadvantages?
A. I. W. T. June 1937.

7. How can one estimate the maximum useful frequency generated by a television transmitting system? What assumptions have to be made? *A. I. W. T. June 1937.*

8. Why does a television subject have to be scanned? Discuss the relationship between the number of scanning lines viewing distance and picture size. *A. I. W. T. June 1937.*

9. What are the basic factors which combine to make up a scene? Discuss the possibility of transmitting each of these by television.

10. Describe a television system in which sequential scanning is not employed. What are the advantages and difficulties of such a system, and why is it not in use to-day?

11. Show why a lensed disk is not more efficient than a perforated or Nipkow disk when used for high-definition television.

12. How is the term 'definition' used in television terminology? How may it be defined? What do you regard as typical examples of high and low definition, and what sort of scenes would you expect to transmit adequately in each case?

CHAPTER XVI

PRINCIPLES OF MODERN TELEVISION WORKING

THE use of a cathode-ray tube for vision reception was probably first proposed by Rosing in 1907. In 1908 Campbell Swinton described a system having much in common with those now used, suggesting for the first time cathode-ray tubes at both transmitter and receiver. Whilst these workers share the credit for visualizing the advantages of an inertialess beam, the crude Braun tubes of that date were far too undeveloped to permit of any practical application of these ideas. Many subsequent workers shared these views, and carried out research with the best equipment then available, but for a number of years cathode-ray technique was inadequate to meet their needs. It is in fact very largely the demand by television workers for a satisfactory tube, and their contributions to the subject, that has led to the efficient tubes now available.

We have noted the underlying principles of the cathode-ray tube in Chapter V, and have studied its construction and development. It was noted that early tubes employing a strip filament with perhaps a small area of emitting oxide coating, often had only a few hours life. This was particularly so in many of the gas-focused types, in which electronic bombardment quickly destroyed the filament.

The original Braun tube and those based on it were high vacuum types, needing very high anode voltages, and most inconvenient to use. The gas-focused tubes removed this defect and would work well with from 400 to 1,000 volts. They were also more sensitive, and began to achieve real popularity in the laboratory; but television workers found that the focusing method failed to yield a sharp trace at high scanning speeds, thus destroying the chief advantages of the method. As recently as 1925 tubes were unreliable and suffered from most of these defects, whilst they were built without any means of modulating the beam intensity. This could not be done without upsetting the focus. Thus cathode-ray television was compelled to wait upon tube development.

By a process of gradual improvement cathode-ray tubes are now available free from these defects, and having characteristics

suited to television. They are built by methods similar to those used in the mass production of valves, and their cathodes have a similar life. They are also approaching the same standards of reliability in other respects, and replacements are interchangeable. These tubes are of the high vacuum electron focused variety, having an electrode structure similar to that of Fig. 45.

Other examples employ the effects of a magnetic field set up by a solenoid concentric with the axis of the beam, and which has the effect of constricting the electron stream. Under the influence of such a field, adjusted to an optimum intensity, the electrons follow helical paths and converge to a common point on the axis, at which the screen is situated. The negative cylinder and first anode are still necessary in this type, to form and accelerate the beam, but the remaining electrodes are replaced by a simpler magnetic arrangement. We have already seen that the beam can also be deflected by either electrostatic or magnetic action, and it is a matter of convenience which is employed. Magnetic deflexion cheapens tube construction, but may have the reverse effect upon the associated circuits.

A television image received on the cathode-ray tube is of course built up from a fluorescent glow set up upon the screen by the impact of the electron beam. The latter is the scanning agent, and the spot of light which it forms on the screen furnishes the scanning 'aperture' or element. The beam must therefore trace out a scanning pattern or raster, and must move in a series of parallel lines. Each line will be of the width of the spot, which must be capable of accurate focus to a sufficiently small size, and must remain focused whilst modulated.

Early tubes were unsatisfactory in this respect, and whilst the problem of focus has been largely overcome, there is still a limit to the size of spot obtainable. Now for an image built up of N scanning lines the total width of the field will be approximately N times that of a single line, namely, of the spot itself. Hence there is a limit to the number of lines which can be accommodated upon a screen of any given size. An attempt to compress more lines into a limited screen area will result in overlapping between them and loss of detail. This factor makes it very difficult to produce a sharp image of small size, and is a difficulty in the way of the projection of a small bright image on to a larger screen. Similarly, the spot must be enlarged if an

image is to fill a larger screen, so that the lines will fill the screen completely and will not leave dark strips between them.

A usual size for the screen of laboratory tubes has been from 4 to 6 inches in diameter. Screens are normally circular, a rectangular area being masked off to frame the image when thought necessary. They cannot usually be quite flat, as it is found that only a curved bulb can withstand the atmospheric pressure when evacuated. A spherical surface possesses this essential strength, and therefore the screen will be a section of an approximately spherical surface, of the smallest curvature found to be safe. A slight risk of collapse still exists, started perhaps by a slight flaw in the glass, and since this will result in the electrode assembly being projected through the screen end of the tube with dangerous force, it is advisable to view the screen through a window of safety glass.

This slight curvature of the screen prevents a satisfactory image being obtained near its edges, and reduces the useful area somewhat. It is also used as an argument in favour of optical-mechanical methods by those who prefer these. Tubes are now made with screens of all sizes from 1-inch diameter upwards, and in the smallest sizes the screen can be quite flat. For high-definition television the minimum satisfactory size is about 8 inches, an average size from 10 to 14 inches, whilst a few larger tubes of 16 inches and more are being developed. We may take about one square foot, however, as typical of an average image at the present time.

Many attempts have been made to magnify and project cathode-ray images on to a large flat screen. Definition seems to suffer by optical projection, possibly owing to the curved screen and to the fact that the glow possesses slight depth, whilst it is not easy to obtain a bright enough image to allow of considerable magnification. Very recently, however, good results have been shown by several firms, and it seems that larger images may soon be expected by projection from specially designed tubes.

The spot produced by a cathode ray is roughly circular in shape, and surrounded by a slight 'halo' or halation. It is thus less sharply defined than those obtainable by devices such as the scanning disk, and cannot readily be made rectangular or hexagonal. For this reason it can be said that for a given number of lines the cathode ray will not give quite such good

image sharpness as the mechanical systems. This fact is offset by the ease with which it will deal with far larger numbers of lines, whilst research is leading to the possibility of a sharper element having definite shape.

The minor defects mentioned in images at present obtainable by the cathode ray become unimportant in comparison with the enormously improved brightness which it offers. This is one of the two factors which make high definition possible by its use. We have noted that the image is produced by a fluorescent glow, and this is set up by the impact of the beam. The brightness of this glow depends upon the nature and purity of the screen material. It is also proportional to the energy contained in the beam, which is in turn drawn from the tube anode current. We can increase this energy almost indefinitely by the use of high-anode potentials. There is thus no serious limitation such as that imposed by the poor optical efficiency of mechanical scanning, and images can be viewed in a moderately lighted room.

The improved illumination which makes this possible is explained by the different laws governing the illumination of the cathode-ray screen in comparison with those governing a beam of light. We have seen that the latter must necessarily suffer from high-scanning speeds, and that a screen scanned by a beam of light must inevitably fall off in average illumination as the square, or an even higher power, of the number of lines used. The cathode ray, however, imparts its energy to the fluorescent material with little regard to its rate of motion. Thus the total energy and total screen illumination for a given beam energy tends to remain constant at all scanning speeds or numbers of lines. It makes very little difference if the beam scans the screen slowly in a few lines or rapidly in a large number. Provided that the whole area is covered in a given period, the energy received by the fluorescent material will be the same, and will set up the same average illumination. Thus definition can be increased without any marked loss of image brightness, although practically a small and unimportant reduction does take place.

The second factor in favour of cathode-ray working lies in the negligible inertia of the beam. This will allow it to be deflected at high velocity with ease and to scan any number of lines likely to be required in practice. A limiting factor exists

in that the beam must remain sharply focused at these high velocities, but it will do so in the case of well-designed high-vacuum tubes employing electron focusing.

The tube is silent in operation, free from moving parts, and conveniently operated by purely electrical methods, facts which still further increase its attractiveness, with the result that no alternative methods seem equally suited to domestic television reception. Whereas mechanical systems have been found cumbersome at even 100-line definition in the past, and have not been used with success above 240 lines in commercial equipment, the cathode-ray tube has enabled a standard of 405 lines to be adopted for the London broadcasting service. It would seem possible to double this number if other considerations allowed.

Cathode-Ray Circuits. We must now turn briefly to the circuits associated with the cathode-ray receiver, which are of equal importance to the tube itself. Firstly, there will be supply potentials for the various electrodes, and whilst these might be derived from batteries in the laboratory, an eliminator unit supplied by the alternating-current mains is the rule in nearly all commercial equipment.

This is designed on the lines already explained in Chapter VII. A typical tube may require a maximum anode potential of from 2,000 to 3,000 volts, but fortunately at a fraction of a milli-ampere only. This demands a well-insulated mains transformer of the desired rating, followed by a half-wave valve or copper-oxide rectifier designed to withstand high peak potentials. A simple smoothing filter will suffice, in which the usual choke can be replaced by a resistance of perhaps 250,000 ohms. At such a small current the potential drop across this will not be serious. A capacity of 1 mfd. or less is likely to be sufficient for each smoothing condenser. The resulting voltage is directly applied between the main anode and tube cathode, the former being usually the earthed point of the system. A cathode-heating transformer insulated to withstand the full anode potential is therefore also necessary. The lower positive potentials needed for the remaining 'anodes' which make up the electronic focusing system are derived from a potential-dividing network, the arms of which are adjustable to the extent necessary for focusing.

The negative cylinder determines the beam current to a large extent, and will require an exactly adjusted bias. This can be derived across a resistance through which the beam current

flows, by analogy with valve cathode biasing, or from one of the other eliminators, such as that supplying the time bases. It is treated as an adjustment for the mean screen brightness, and is a front-panel control in most receivers.

The beam intensity, or current, must be modulated to provide light and shade in the reproduced image. This corresponds to the light valve of an optical system. In early tubes it was effected by modulating the negative cylinder potential by the vision signals, but since this potential influenced the focus, dark areas were partly produced by defocusing and dispersing the beam. As a result, bright portions of the image would be sharp and detailed, but the darker portions became blurred. This defect applied particularly to gas-focused tubes.

Improved electrode design has ensured that negative cylinder potential controls merely the number of electrons passing through the first anode aperture, focusing being entirely electronic and carried out at a later point in the beam's path. It has recently been decided to refer to the negative or Whenelt cylinder as the 'grid', since it can be regarded as controlling anode current very much as does the grid of a valve. This term will be used for any other electrode employed for that purpose in tubes of different electrode design. At one time the anodes were termed 'accelerators', but it has been decided to give up all such terms which sprang up in the earlier days of development in favour of 'first, second, or third anodes', the first anode being that next the cathode.

Modulation can be applied to the grid through a simple resistance-capacity filter, a condenser supplying the modulating potentials, and a resistance passing the direct biasing potential. In recent television systems, however, the mean image brightness is transmitted in the form of the mean amplitude of a carrier wave, and it is found best to rectify this by a diode and to apply the whole product to the tube grid directly. The mean rectified carrier then controls the mean grid bias, and determines the mean image brightness, so that the general tone of an image can be transmitted. We term this utilizing the 'D.C. Component' of the transmission. It ensures that a scene in full sunlight, for example, is reproduced more brightly than one on a dark day; whilst if the D.C. bias be suppressed all scenes become of the same average brightness.

To complete the receiving process it remains necessary to

deflect the beam in such a manner that it reproduces the scanning raster of the transmitter. We desire the beam to travel across the screen with strictly uniform velocity, after which it must return to the origin in negligible time and commence the tracing of another line slightly displaced from but parallel to the first. This process will be repeated until the required number of lines have been traced, when the beam must return quickly to its starting-point and commence the raster afresh.

It will be seen that this motion can be produced by two oscillators, one of which deflects the beam along the lines at the 'line frequency' whilst the other deflects the lines slowly across their length, completing an oscillation in the time period of one raster, namely, at the 'frame frequency'. These two motions correspond to those of the line and frame scanners of a crossed mechanical system. The deflecting potentials must not be sinusoidal, however, but 'saw toothed' in form, as shown in Fig. 170 (*b*). For the duration *AB* of one line (or frame) the potential must increase linearly, and at the correct instant *B* must return as quickly as possible to its initial value *C*. Circuits to produce this wave form are termed time-base generators, or sometimes saw-tooth oscillators. They are of many types and cannot be reviewed in detail, but we shall outline a simple form that is widely used. It may be added at this point that magnetically focused or deflected tubes are treated very similarly, but that currents will be employed wherever we have referred to potentials.

Fig. 170 (*a*) shows the essentials of a time-base generator of the type employing a thyratron or gas-filled triode. It is flexible in adjustment and ideal for laboratory purposes. The deflecting potential is set up across the condenser *C*. This should be charged through a constant-current device in the position *R*, for if a constant current flows into the condenser, the potential across it will rise linearly. A diode valve was amongst the first devices used in place of *R*, operated at a low filament temperature which results in saturated emission, and a low constant anode current independent of potential. This scheme is critically dependent upon filament temperature, and is somewhat unreliable.

A better charging device is the pentode, which has an inherently constant-current characteristic between certain limits of anode voltage. The charging rate is determined by grid bias,

being thus readily adjustable, and the method is very satisfactory. A cheaper and simpler method, however, merely uses a high resistance R, of several megohms, and adjustable to determine the rate of charging. The potential across C will now rise exponentially, which is unsatisfactory; but it so happens that the lower portion of an exponential curve (the first tenth) is

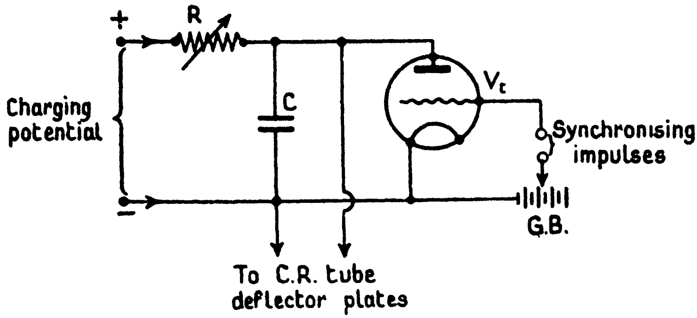


FIG. 170 (*a*). Simple Time-base Oscillator.

sufficiently linear for tolerable television results. The method demands a high charging potential of the order of 1,000 volts, whilst the potential across C is not allowed to rise by more than some 100 volts, thus keeping to the linear portion of the curve.

C and R must be adjusted so that this potential is reached at the end of the time required to scan one line, namely, at B of Fig. 170 (*b*). At this point the gas-filled valve V_t discharges the condenser rapidly. Such a valve suddenly becomes conductive when the potential across it reaches a predetermined value depending upon the grid bias, when ionization sets in. Its anode resistance remains low as long as there is any considerable potential across it, after which ionization ceases. The grid bias can thus be adjusted so that discharge occurs at regular intervals, such as the period AB, when the critical potential is reached across C. It can be assisted to discharge at a given instant if an external synchronizing potential be added to this bias.

The modern vision signal is provided with a synchronizing impulse occurring at the exact end of each scanning line. If this be applied to the thyratron grid, it can momentarily reduce the grid potential. If the time constant of R and C has been adjusted so that the discharging potential is nearly reached at that instant, a reduction of bias will 'trigger' the thyratron,

terminating each line at an exact instant corresponding to the transmission. In this way cathode-ray television reception can be synchronized perfectly by the vision signals themselves.

The frame or low-frequency time base is similarly designed. It is synchronized by a similar impulse occurring at the end of

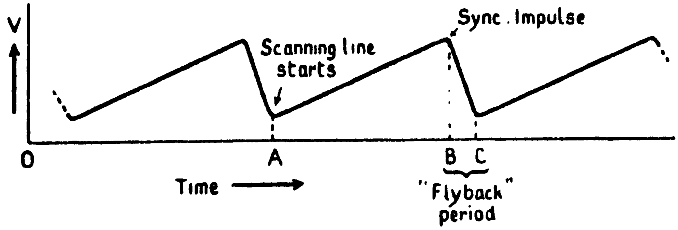


FIG. 170 (*b*).

each frame raster, and of longer duration so that it can be separated from the line synchronizing impulses by a circuit of longer time constant. A thyratron containing mercury vapour is quite adequate for the frame time base, but for that furnishing the line scanning it will not discharge C sufficiently rapidly. A filling of argon or helium is preferred in this case. A small resistance may be inserted in series with the anode circuit to limit the peak discharge current to a value safe for the valve.

Whilst generally satisfactory, time bases depending upon the gas-filled triode are said to be slightly irregular in action at high speeds, owing to variations in the discharge rate. The discharge period, or 'fly-back period' BC, is also not as brief as might be desired, reaching perhaps 10 per cent. of the line period at high-definition speeds. To overcome these defects various other forms of oscillator have been developed in which only 'hard' valves are used. In one type a regenerative oscillator is used, in which a special iron-cored transformer provides coupling between grid and anode. The core of this works at or near magnetic saturation, resulting in a distorted wave form which can be made to approach the saw-tooth. In other 'relaxation oscillators' a pair of valves operate very much in the manner of a multi-vibrator, which we saw can produce a wave form of the desired type.

Beam deflexion is not without possibilities of distortion, and it is found that unless each pair of deflector plates are kept at a symmetrical potential with respect to the focusing system,

the rectangular raster will be drawn out into a trapezium. This 'trapezium distortion' can be prevented if deflecting potentials of equal magnitude and opposite phase be applied to the plates, so that as one of a pair becomes more positive the other becomes more negative. They should, in fact, be excited in push-pull. In addition, the potential necessary to deflect the beam fully in a large tube will exceed the output possible from most types of time-base generator, reaching several hundred volts.

Both these difficulties can be overcome by the use of a two-valve amplifying arrangement, in which one valve amplifies the generator output for application to one deflector plate, whilst the other acts as a phase reverser to supply an equal potential in opposite phase to the other plate. Special triode valves capable of amplifying large potentials are made for such work. Slight distortion introduced when the time-base condenser C is charged exponentially can be corrected in this amplifier, equal and opposite curvature being produced by the valve characteristics. It is usual to combine the line and frame time bases together with their amplifying valves into a single unit when designing commercial television receivers. Certain subsidiary controls are also needed, such as those for the amplitude applied to each pair of plates, by which the size and shape of the raster can be adjusted correctly; and potentiometers which provide a steady bias between each pair by which the image can be 'framed' or centred on the screen.

Television Cameras. Before considering other aspects of television reception, it will be best to see how modern high-definition images are transmitted and the type of vision signal which results. Mechanical-scanning systems of the types described in the previous chapter are perhaps better suited to transmission than to reception, for complex equipment is more acceptable at the transmitting station. Preliminary tests during the experimental period at the Alexandra Palace used several of these. In one type of transmitter developed by the Baird Company intermediate film was employed. Mechanical film scanning is a perfectly satisfactory process, since ample light can be used in the projector. 240-line images at 25 frames per second were transmitted successfully, the scanning by this process being very accurate. For studio scenes a film was 'shot' in the ordinary way, quickly developed and fixed within a few seconds, and passed through a similar disk scanner for vision

transmission. Intermediate film working was found inconvenient, however, in a variety of ways, being poorly adapted to programme production.

An alternative method has been found in the electronic camera, developed along different lines by both the Baird and E.M.I. organizations. The Baird electron camera employs the

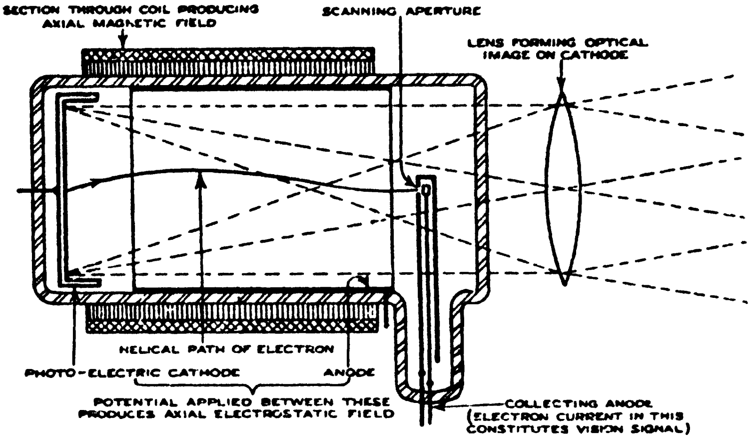


FIG. 171. Baird-Farnsworth Camera.

principle illustrated in Fig. 171. The lens or photographic objective forms an image of the illuminated scene upon a plane photo-electric cathode, with the result that electrons are emitted from each portion of the surface in proportion to the illumination falling upon it. An axial magnetic field is set up by a solenoid surrounding the tube, with the result that these electrons are brought to a focus in the plane of the scanning aperture, to form an 'electron image' of the scene in that plane.

This effect is obtained by the technique evolved for the magnetic focusing of a cathode ray, each portion of the cathode being regarded as emitting its own beam of electrons, which follow similar helical paths. In this plane is also a small collecting electrode, screened except for a small scanning aperture through which an element of the electron image can pass, to produce a photo-current. Scanning is effected by the reverse of the usual process, the whole electron image being swept across the collecting electrode by the external field exactly as in the cathode-ray tube. This is clearly equivalent to scanning a fixed image by a moving element in the more conventional way.

The sensitivity of this device is not fundamentally superior to that of a normal photo-cell, and to make it effective at high definitions reliance is placed upon the electron-multiplier tubes mentioned in Chapter V. By these the weak current from the collector electrode can be magnified without the limitations of 'valve noise'. The minimum input which can be amplified may be about 100 times less than would be possible if it were taken to the grid of a hot cathode valve. Scanning speeds, of course, present no problems since only an electron beam has to be deflected, and the same advantages over mechanical scanning exist as in cathode-ray reception.

The form of electron-scanning tube used by E.M.I. is perhaps of more fundamental importance, because it gives for the first time a large inherent increase in sensitivity over mechanical scanning. The tube lends itself to use in a camera housing very similar to a film camera and similarly handled, whence it is termed the Emitron Camera. Here a lens combination P forms a real image upon a photo-sensitive cathode Q as before, but the structure of this cathode is essentially different. Instead of being a continuous surface, the cathode is built up in mosaic form upon an insulating plate such as mica. This is coated upon the back with a conducting metallic layer, which forms one collecting electrode R in Fig. 172. Upon the front this mica plate is coated with numerous minute globules of silver. These are sensitized with a photo-sensitive material such as caesium oxide by a special chemical process. Each of these globules is insulated from its neighbours and forms an elemental photo-cell, whilst it also forms a small electrical condenser of which the back coating is the other electrode.

Scanning is carried out by a cathode ray, which sweeps the front of the mosaic cathode exactly as it would the fluorescent screen of a receiving tube. The necessary electrode assembly to produce and deflect this beam forms an integral part of the evacuated bulb A in which the mosaic screen is mounted, in the manner shown at C, D, E, F. The scanning beam H plays the part of an inertialess commutator, for as it scans the surface of the mosaic of elemental cells it picks up the electrical charge that has been accumulating upon each cell and transfers it to the amplifiers. The beam can be said to complete an electrical circuit from each elemental area in turn, current passing from the back coating R through the external circuit ST to an

inner conducting coating G upon the tube walls. This in turn provides the return path for the beam current. In this way each small condenser formed by each elemental cell will be discharged as the beam passes over it, the charge reaching an external

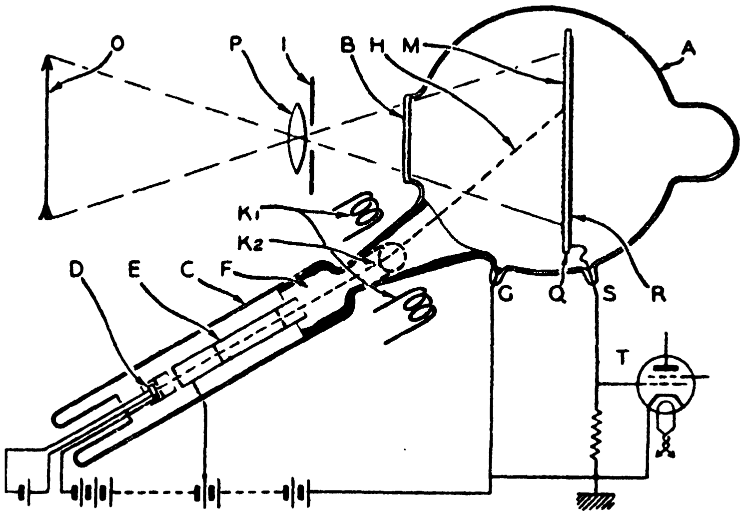


FIG. 172. Marconi-E.M.I. Emitron Camera.

resistance across which is developed a potential that can be amplified and used to modulate the radio-transmitter.

The improved efficiency of an Emitron tube arises from the fact that all light collected by the lens P is acting continuously upon the photo-sensitive surface. Substantially the whole of it produces emission, which is stored up in the form of an electric charge upon the elemental cells. At each occasion that the scanning beam passes over any one of these, it collects most of the energy which has accumulated there since its last passage, and in this manner most of the light reflected from a scene gives rise to useful photo-electric current. There is no inherent loss of energy due to the scanning process, for the beam does not merely collect the charge formed by each cell during the instant in which it passes over it, but that which is continuously accumulating at that cell.

The efficiency of early mosaic systems has thus been largely recaptured, but without the defects of such devices as mechanical commutators, multiple channels, lamp screens, and the like.

There is, of course, some loss in the Emitron system, but this is due to technical or constructional imperfections rather than to basic causes which cannot be overcome. There seems no reason why it should not be developed to deal with still higher definitions than 405 lines, whilst at this standard good results are possible from scenes illuminated only by weak daylight.

The London Vision Transmitter. No matter what system of scanning or type of television camera is used, the photo-currents produced are likely to be very minute. The potentials which they set up across an input impedance must therefore be amplified, and to do this with the least risk of external noise pick-up, a 'head amplifier' is used as near to the photo-cells as possible. In the Marconi-E.M.I. equipment as used by the B.B.C. a four-stage amplifier is incorporated in the camera itself, being thus very close to the Emitron tube and in the same screened compartment with it. This unit raises the vision signals to a level at which they can be safely passed over a moderate length of cable to the main amplifiers. Owing to the high-maximum frequencies involved, the cable must be of exceptionally low capacity, and is made up in the form of a multi-core screened conductor which also carries the operating and deflecting potentials to the tube.

A number of cameras may be used in convenient locations inside or outside the studios, and are connected by cable lines to separate 'A' amplifiers. Here are incorporated controls and circuits akin to automatic volume control, which compensate for irregular changes of illumination or unevenness to which the cameras are subject. They are followed by a circuit for the reversal of phase. This ensures that the signals finally sent out are always of the phase decided upon as 'positive', even though it may be convenient to transmit from a film negative on some occasions.

Signals from the various cameras in use having been equalized in amplitude and phase, can now be taken to mixer panels remotely controlled from the producer's desk, and combined or interchanged as the programme demands. They can be combined in any desired groups, and the signals from one group faded into those from another, very much as in the sound-recording studio. From the mixing panels only two lines are found necessary, each of which passes to a 'B' amplifier. One of these two 'B' amplifiers will normally be in use for the

radiated programme, whilst the other provides a spare against emergency break-down and furnishes a separate signal to monitor receivers, on which the scene picked up by the other cameras can be inspected before it is faded into the main transmission.

The 'B' amplifiers are each of four stages, and include filters for the reduction of unwanted impulses which tend to arise during the synchronizing periods at the end of each line and frame. They include controls for amplification, and are followed by a pair of 'C' amplifiers, each of three stages.

The signal amplitude is now sufficient to feed the final amplifiers which modulate the radio transmitter, but before doing so the special synchronizing impulses must be added. The signals pass through panels whose duty it is to complete the removal of unwanted impulses which arise in the cameras at the end of each line or frame, on account of the interruption of the scanning beam found advisable at those instants. A portion of the signal is entirely removed to provide the spaces into which the synchronizing impulses will fit, after which it passes to a contrast control panel by which any imperfections in the balance of light and shade can be corrected. The next stage is the insertion of the synchronizing impulses themselves, after which the signals pass to the main modulation amplifier, and in addition to a number of separate monitoring circuits.

The elaborate equipment needed to transmit high-definition television is not completed, however, with this series of amplifiers and correcting circuits, for there is also a group of panels required to generate the various scanning and exciting potentials. The main scanning frequency is produced by a master oscillator at 20,250 cycles per second, which is synchronized by the 50-cycle alternating-current mains supply. This oscillation has a wave form rich in harmonics, two of which at the line frequency of 10,125 cycles and the frame frequency of 50 cycles are selected by subsequent circuits and passed to the synchronizing signal generators.

The frequencies of 20,250, 10,125, and 50 cycles per second now available perform a number of duties. Suitable 'generators' produce a rectangular wave form which is fed to the cameras for the purpose of wiping out the beam during the flyback periods of scanning. Another sets up the necessary saw-toothed scanning potentials, synchronized by the 10,125- and 50-cycle impulses. A third feeds correcting impulses to the 'A' amplifiers,

where they are applied in reverse phase to cancel the unwanted impulses already referred to from the cameras. Our description of these many functions is brief, for they are only necessary to counteract the particular defects of the particular system of transmission in use, and form no part of the general study of television principles. That they are fully justified is well shown by the remarkable quality and freedom from distortion to be seen in the resulting images.

The final distribution of television signals is possible either over a cable or a radio channel. In the latter case a frequency must be selected at which a two-megacycle sideband width can be accommodated, as has been discussed earlier. It must be free from fading and distortion caused by mixing of the ground and reflected waves. A frequency of 45 megacycles has been selected for the vision-carrier wave of the London Television Service, and one of 41·5 megacycles for the associated sound transmission. These frequencies correspond to wavelengths slightly over 7 metres, which are in the ultra-short region, and obey quasi-optical laws of propagation. To obtain a good service area for the direct waves it is necessary to erect the transmitter on high ground, having an extensive horizon.

Only a few sites near London meet this requirement, and after taking many others into consideration the old Alexandra Palace was chosen as providing suitable accommodation and facilities. Situated in North London, only 6 miles north of Charing Cross, the ground rises to some 300 feet above sea-level. In addition, a square-section tapering steel tower exceeding 300 feet in height has been erected upon one of the original palace towers, giving a total height exceeding 600 feet. The peak power output of the vision transmitter is some 17 Kw., and this produces a field strength of about 1 millivolt per metre at a distance of 20 miles, if the receiving aerial is clear of the ground and screening objects by some 30 feet. The service area can be taken to have a radius of quite 30 miles, whilst good reception at several times this distance is probable in certain directions.

Modulation of the transmitter is designed to retain the direct-current component of the photo-cell output, which corresponds to the average brightness of a scene. Whilst this is not absolutely essential, since detailed images are possible without it, its use leads to improved reliability and definite modulation levels,

whilst it ensures that the mean brightness of the received images shall correspond to that of the transmitted scene. Clearly it would be most difficult to amplify the direct photo-current through the main system, which is only designed to deal with potential fluctuations or the alternating component. It is therefore better to pass the direct-current component through

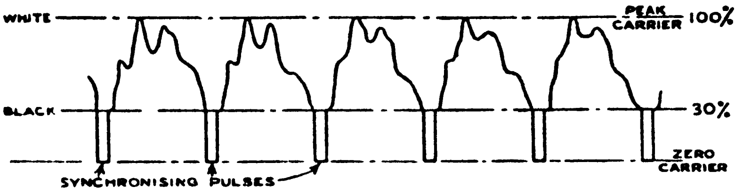


FIG. 173. Wave-form of Television Transmission.

a separate channel. Since it is relatively large it can be amplified by a simple direct coupled amplifier and added to the transmitter modulation circuits later. The latter employs a Franklin-type master oscillator at half the radiated frequency, followed by a frequency doubler stage and five stages of signal frequency amplification. The final amplifier is modulated by a special form of grid bias modulation, which can handle the wide frequency band needed.

Fig. 173 illustrates the wave form radiated after all these steps have been completed, and will make some of the foregoing explanations clearer. It is arranged that a modulation depth (carrier amplitude) of 30 per cent. of the maximum shall represent complete darkness of the image. Carrier amplitudes above this represent brighter portions of the image, 100 per cent. modulation corresponding to the brightest areas, or 'full white'. The carrier amplitude above 30 per cent. therefore corresponds to intensity of illumination at any instant, and is controlled both by the direct-current and alternating components of the photo-electric potentials. Below 30 per cent. carrier amplitude no further reduction in illumination is possible, and this region is said to correspond to a 'blacker than black' condition, which is reserved exclusively for the synchronizing impulses. During those periods the carrier ceases completely, a condition produced partly by the suppression of the image modulation, and in addition by the application of a negative modulation pulse of rectangular wave form. Since all image modulation occurs in the region above 30 per cent., and the synchronizing impulses

entirely below this level, the fact can be made use of to separate them in the receiver. This is done in the following way.

A fixed bias is applied to the receiving cathode-ray tube, which just reduces the beam to zero when a carrier amplitude of 30 per cent. is being received. Any deeper modulation produces increased positive bias and produces illumination of the screen proportional to the vision signals; but a reduction below 30 per cent. during the synchronizing pulses cannot reduce the screen brightness below zero, and so has no effect upon the image. The received signals are also applied to a filter which will select the synchronizing pulses without any mixture of vision modulation.

Such an amplitude filter is provided by a saturated pentode, working with a limited anode and screen potential. If the anode current of this valve is already at its maximum saturation value when 30 per cent. rectified carrier is applied to its grid, no increase in positive bias during the vision modulation can affect it. On the other hand, a reduction of carrier below 30 per cent. reduces the grid bias, and consequently the anode current, which shows a sudden drop during the synchronizing pulses only. These pulses can be transferred to the thyratron grids or other control points of the time-base generators, which are thus 'triggered' to finish each line at the instant determined by the corresponding generators at the transmitting station.

We have now described all the more important features of a modern television transmitter, except the actual radiators themselves. Since it is desired to radiate as much power as possible in the surface wave, the aerial system should favour low-angled radiation. Reflected waves are unlikely to be useful in vision reception and represent wasted power, whilst they may even cause echo effects and distortion. Horizontally polarized ultra-short waves are known to be the most effective over indirect paths, and it is therefore natural to employ vertically polarized waves in this case. The aerial is for this reason built up of an array of vertical half-wave radiators. A number of these are used, eight being placed in a circle around the aerial tower, with a second similar ring above them to handle the sound programme. The use of a number of radiators has been shown to increase the radiated power for a given total input, whilst it also assists uniform radiation in all directions, and reduces high-angle radiation. To increase the outward radiation and reduce

losses in the tower, a half-wave reflector is placed behind each radiator, thus adding to the beam any power which might otherwise be absorbed by the supporting structure.

Power from the transmitters is taken over low-impedance concentric-type feeders, which terminate near the aerials in impedance-matching transformers T. These step-up the low-

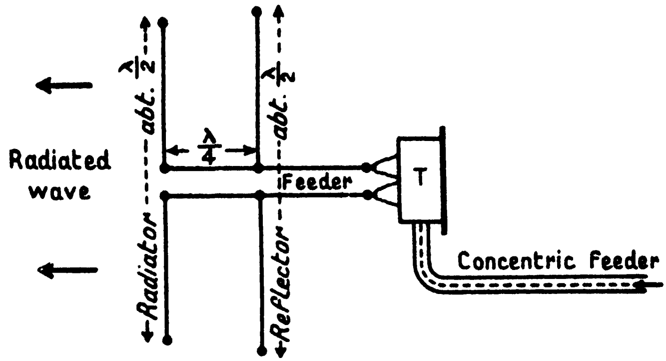


FIG. 174. Vision Aerials.

feeder impedance to a value which matches that of the radiators, which are centre fed by short resonant feeders, whilst the same feeders also excite the reflectors in the appropriate phase lead of 90°. This is ensured because each reflector is a quarter wavelength behind the corresponding radiator, as shown in the single pair illustrated in Fig. 174. The vision aerial is situated at the highest point, whilst below it a similar ring of radiators is used for the sound transmission.

The receiving aerial will be designed to receive vertically polarized waves, but it is usually unnecessary to employ a separate system for both sound and vision. A half-wave aerial is chosen to resonate either at the vision-carrier frequency, or some frequency intermediate between this and the sound carrier, which can be selected so that both transmissions are received at similar strength. The single half-wave aerial will be mounted vertically, as high as possible above surrounding objects. Where field strength is poor it will be an advantage to add a half-wave reflector, placed a quarter wavelength from the aerial on the side away from the transmitter. This will be the usual 2 to 3 per cent. longer than the former. and need not be connected to the feeder in any way.

Some form of non-absorbing feeder is very desirable to couple the aerial to the receiver, since interference from motor-car ignition systems and domestic or commercial electrical apparatus may be picked up on an open feeder. The low-impedance types in the order of 80 ohms are almost universally used, the actual cable being either a parallel or twisted pair preferably insulated by a special low-loss compound suited to this work, or one of the proprietary concentric feeders. This may be inserted directly into the centre of the aerial where a reasonable impedance match exists, or coupled by one of the methods studied in Chapter XIV, such as that of Fig. 162.

Radio receivers for the vision signal fall into two classes, the tuned radio-frequency amplifier, and the superheterodyne. The former may incorporate four or five amplifying stages, which must be tuned sufficiently flatly to accept the wide sideband radiated. A degree of damping will be introduced for this purpose, and it is very helpful to 'stagger' the resonant frequency of the couplings somewhat. This results in a wider band-pass response curve with less loss of amplification than would occur if damping only were relied upon. Tuned anode coupling is suitable, there being no advantage in the increased number of selective circuits yielded by transformers.

Pentode valves of high mutual conductance, combined with fairly low impedance and minimum inter-electrode capacities will be selected, but in spite of this their anode impedance will be so undermatched that a very poor stage gain is inevitable. This is the reason why so many stages are necessary. As an example, the E.M.I. receiver employs five stages, followed by a diode detector which directly feeds the cathode-ray tube. Sufficient sensitivity for all normal requirements is claimed, together with a minimum of distortion and interference. Very complete screening is naturally necessary to stabilize such an amplifier at very high frequencies.

The superheterodyne is possibly a little less perfect in these respects, but is simpler to design and more widely used. It may include a single radio-frequency stage, followed by a triode-pentode type of frequency changer. The intermediate frequency must be high enough to deal with a 2-megacycle sideband, but not too high for reasonable stage gain, whilst it must be chosen to avoid possible interference with harmonics of the sound transmission, and like difficulties. The usual value lies between 5 and

10 megacycles. At this frequency the design problem is not very different from that of a signal frequency amplifier at 45 megacycles, for whilst a better coupling efficiency is to be expected at the lower frequency, this is largely offset by the higher damping necessary to flatten the response curve adequately.

An advantage can be gained if single-sideband reception is used, when the amplifier need only accept a total band of 2 megacycles as against a total of 4 megacycles when both sidebands are amplified. In any case, a stage gain of 16 seems about the maximum, whilst lower figures are general, and in consequence from three to five intermediate-frequency stages are employed before the diode detector. The latter will be a special low-impedance type, which works into a load resistance of some 5,000 ohms and a parallel capacity of about 15 mmfd. in order to retain the higher frequencies.

In some designs reduced intermediate-frequency amplification is made up by the use of an amplifying stage following the diode, which may add a gain of about 20 at the expense of slight distortion. This is termed a vision frequency or V.F. amplifier, by analogy with an L.F. stage. It may be direct-coupled to retain the D.C. component of the vision signals, or may be resistance-capacity coupled. In the latter case a third rectifier is necessary after the V.F. stage, in order to provide a bias proportional to the mean carrier voltage, and thus to restore the lost D.C. component at the cathode-ray tube grid. A metal rectifier can be used for the purpose.

The receiver for the accompanying sound may be an entirely separate unit of conventional design, a simple superheterodyne being favoured. No exceptional number of stages is necessary in this case, but it has been found possible to transmit a full frequency modulation up to at least 10,000 cycles per second, and it is thus advisable to provide for this range in the receiver also. As a result, the quality of reproduction is materially better than the demands of selectivity allow at lower broadcasting frequencies. It seems highly probable that increased use will be made of the ultra-short wavelengths for high-quality local broadcasting in the near future. The wide spectrum available should allow of a 20 Kc. spacing between transmissions, with a range of reproduction practically up to the audible limit.

In many designs the early stages of the vision receiver are sufficiently flatly tuned to accept the sound signals also. This

fact may be used to simplify the sound receiver, which may derive its input from an intermediate point in the vision circuits. It is possible to employ common frequency-changing circuits for both, a separate intermediate-frequency amplifier only being provided at the appropriate sound frequency.

In concluding this brief outline of television methods it may be added that transmission is not only possible by radio, but that cable links are also practicable if the electrical capacity be sufficiently reduced. The present tendency in communication by cable is to replace existing circuits by new cables which can deal with frequency bands up to several megacycles. By the use of sub-carrier or 'wired wireless' methods it is possible to handle a large number of telegraphic or telephonic signals simultaneously over a single high-frequency cable. Since this will be run underground and heavily protected it may be more reliable than overhead wires, whilst in spite of its high cost of approximately £1,000 per mile, the large number of older circuits which it can replace may result in an eventual economy. The British Post Office are at present installing several such cables on such trunk lines as London to Birmingham. They will also be suitable for television transmission. Cables already exist between Broadcasting House and the television station, whilst there is a loop laid round central London to assist in outside television broadcasts.

Two types of cables are being developed by different manufacturers. The one is a concentric cable, in which a thin inner conductor is concentrically spaced by low-loss insulators along the axis of a metal tubular outer sheath which forms the second conductor. It thus resembles a concentric feeder in construction. The dielectric is mainly air, whilst the whole is, of course, heavily protected by outer coverings against the weather. In the second system both conductors are supported parallel to each other along the axis of a tubular outer sheath which forms merely an earthed shield. In this method the advantages of balanced operation exist since neither conductor need be earthed.

Cables of these types, however built, are provided with 'boosting stations' every few miles, at which a valve amplifier raises the signal level whilst introducing correction for the unequal attenuation of higher frequencies. Television signals can be transmitted quite successfully over these improved

cables, which open up the possibility of distant outside broadcasts and relayed programmes. By their use it should be possible to erect a number of vision radio transmitters supplied with a common programme, which will thus be available over a large area.

An alternative exists for short distances in the form of a radio link. Television picked up by a camera, amplifiers, and low-power radio transmitter near some outside news or sporting event can be radiated for reception at the main television station, where the received signals provide modulation for the main transmitter. A wavelength different from the latter is, of course, necessary, and successful relays have been carried out by the B.B.C. using one of 5 metres. The whole equipment is housed in a van together with the necessary generators and supplies for the vision camera. It is not possible to anticipate whether radio or cable links will prove most satisfactory, and it is probable that both will become usual features in television communication.

EXAMINATION QUESTIONS

1. Describe an aerial, and its associated circuits, suitable for ultra-short-wave television transmission.
Institute of Wireless Technology. November 1935.

2. Describe the construction and principle of action of a cathode-ray tube. *I. W. T. November 1935.*

3. Describe briefly the construction and action of the 'iconoscope' (Emitron Camera). Compare its theoretical efficiency with that of a simple Nipkow disk system, giving explanation of any difference.
I. W. T. June 1934.

4. What types of radio receiver are suited to high-definition television signals, and in what chief respects will their design differ from a receiver intended for sound broadcasting?

5. What is the function of a time-base generator in a modern television receiver? Explain the essential action underlying one type of simple time-base circuit.

6. Describe fully, with diagrams, an application of one of the following:

(*a*) A thyratron.
(*b*) A goniometer.
(*c*) A saw-toothed-wave generator. *A. I. W. T. June 1937.*

7. Discuss the various substances that may be used for the fluorescent screen of a cathode-ray tube. *A. I. W. T. June 1937.*

8. Discuss the design of a radio receiver capable of picking up signals within the service area of the London Television Station and operating a cathode-ray tube. Illustrate with circuit diagrams.
A. I. W. T. June 1937.

9. What are the effects of traces of residual gas in the filling of a cathode-ray tube, and of a photo-electric cell? Explain the action upon the electrical performance in each case.

10. What are the difficulties that have been encountered in the projection of television images from the cathode-ray tube on to large screens? Suggest any steps by which they are being overcome.

11. Why is it desirable to transmit the 'direct-current component' of a television signal? What part of the transmitted scene does this represent, and how can it be utilized in the receiver?

12. What is the nature of the wave form transmitted from a modern television station? Explain the introduction, function, and separation of the synchronizing pulses which form part of it.

INDEX

Absorption modulation, 241.
— of waves, 48, 400.
Acceptor circuit, 264.
Adcock system, 440.
Admittance, 30.
Aerials, beam, 450.
— coupling, 332, 463.
— frame, 425.
— harmonic, 446, 459.
— height, 49.
— hertzian, 52, 442.
— length, 445.
— Marconi, 47, 424.
— resonance, 441.
— television, 540.
— 'V' and diamond, 460.
— Windom, 476.
— Zeppelin, 466.
Afterglow, 107.
Aircraft, navigation, 432.
Alexanderson, 7, 122.
Alternating current, 17, 25.
Alternator, 121.
A.C. resistance, anode, 89.
Amplification, 88, 199.
— calculation of, 214.
— direct current, 210.
— factor, 88.
— H.F., 213, 282, 298.
— L.F., 216.
— power, 360, 364.
— television, 536.
Amplitude distortion, 180.
Angle of radiation, 407, 448.
Anode, 71.
— A.C. resistance, 89.
— coupling resistance, 201.
— current, 72.
— dissipation, 91.
— modulation, 236.
— rectification, 174.
— to grid capacity, 92.
Antenna, *see* Aerials.
— effect, 429.
Appleton, 398, 416.
Arc transmission, 118.
Armstrong, 273.
Arrays, 451.
Atmospherics, 414.
Attenuation of waves, 394.
— factor, 49.
Autodyne, 270.
Automatic tuning control, 317.
— volume control, 310.
Ayrton and Perry, 491.

Back E.M.F., 16.
Baffle board, 386.
Baird, J. L., 485, 503.
Band-pass filter, 66, 331.
Band-width, 323, 326, 330.
Barkhausen-Kurz, 149.
Battery eliminators, 184.
Beacons, radio, 433.
Beam aerials, 450.
— valves, 105.
Beat-frequency, 267.
Bellini-Tosi, 430.
Beveridge, 410.
Bornite, 169.
Buffer amplifier, 163.

Cady, 152.
Campbell Swinton, 523.
Capacity, 15, 22.
— earth, 51.
Carborundum, 169.
Carrier wave, 232.
Cathode, 85.
— bias, 286, 374.
— ray tube, 103, 416.
— — television, 519.
Centre of capacity, 49.
Characteristic curves, 73, 76, 88, 94, 363.
Charge, 14, 23.
Choke, control, 236.
— coupling, 206.
— output filter, 366.
Class A, B, and C, 221, 362, 372.
— B power stage, 377.
Clear channel, 329.
Coherer, 3.
Colour, in television, 485.
Colpitts circuit, 137.
Condensers, 22, 58.
— current in, 31.
— electrolytic, 194.
— in parallel, 24.
— in series, 25.
— variable, 296.
Conductance, 51.
— conversion, 304.
Cone speaker, 383.
Continuous waves, 117.
Copper oxide rectifier, 191.
Coulomb, 23.
Coupling, 19, 59.
— coefficient, 62.
— condenser, 210.
— by impedance, 64.

Cross-modulation, 346.
Crystal detector, 166, 261.
— filter, 343.
— oscillator, 156.
Curie, J. and P., 152.
Current, aerial, 442.
— alternating, 17.
— direct, 14.
— in circuits, 28.
— in resistances, 19.
Cylinder, negative, 104.

Damped waves, 5, 56.
Damping, 40.
Decoupling, 193.
Decrement, 40.
Definition, of images, 486.
Deflection, cathode rays, 107.
Delayed A.V.C. 315.
Demodulation, 323, 334.
Detection, 165, 177, 358.
Detector, crystal, 166.
— diode, 74, 180, 358.
— triode, 172.
Dielectric, 23, 42.
Diode valve, 73, 98.
Direct current, 14.
Direct-coupled amplifier, 210.
Direction finding, 427.
Directional aerials, 450.
— microphones, 352.
Disk, scanning, 499.
Dissipation, anode, 91.
Distortion, 350.
— in microphones, 351.
— in receivers, 356.
— in transmission, 356.
Diversity reception, 410.
Driven transmitters, 135, 161.
Duddell arc, 117.
Duddell and Taylor, 398.
Dynamic characteristics, 363.
— resistance, 261.
Dynatron, 96, 160.

Earth connexion, 50.
— screen, 51.
Echoes, 404.
Eckersley and Tremellen, 406.
Eddy currents, 41.
Efficiency, of transmitters, 224.
'E' layer, 402.
Electrolytic condenser, 194.
Electromotive force, 15.
— — applied to circuits, 27.
Electron focusing, 103, 111.
— multiplier, 101.
— optics, 112.
— oscillations, 150.
Element of image, 488.
Elster and Geitel, 502.
Elwell, C. F., 6.
Emission, 70.
Emitron camera, 534.
Energy, 37.
Errors, in D.F., 429, 436.

'F' layer, 402.
Fading, 395, 408.
Farad, 23.
Faraday shield, 473.
Feeder impedance, 468.
— termination, 471.
Feeders, 465.
Field, electric, 43.
— magnetic, 16.
Field-strength, 49.
Filament, 71, 78.
Film intermediate, 509.
— scanning, 508.
Filter, smoothing, 185.
Fleming, Sir A., 73.
Flicker, 495, 519.
Flux, 16.
Fourier, 26.
Frame aerial, 425.
Franklin, 151, 455.
Frequency, 17, 39, 58.
— changer, 294, 301.
— modulation, 249.
— multiplication, 161, 222.
— resonant, or natural, 33.
— speech, 229, 245.
— stabilization, 151.
— telegraphic, 251.
— television, 495.
Full-wave rectifier, 184.

Gain control, 308.
Ganging, 281, 295, 332.
Gas focusing, 111.
Gas-filled triode, 100.
Gettering, 79.
Gill, 149.
Goniometer, 430.
Goyder, 10.
Graphical treatment of A.C., 25.
Grid, 74.
Grid bias, by leak, 131.
— — modulation, 242.
— — of power stages, 362, 373.
Grid condenser, 210.
— current, 75, 131, 139.
Grid-leak detector, 173, 180.
Ground waves, 397.

Half-wave Rectifier, 186.
Hard valve, 75.
Harmonics, 159, 446.
Hartley circuit, 136, 268.
Head telephones, 166.
Heater, 85.

Heaviside, 401.
Henry, 17.
Heptode, 96, 303.
Hertz, 2.
Hertzian resonator, 2.
Heterodyne, 266.
— interference, 339.
— separate, 270.
High-frequency amplifier, 213, 282, 298.
— — current, 14.
— — transformer, 214, 289.
Homing, 435.
Homodyne, 267.
Horn loud speaker, 382.
Hum, filament, 83.

Illumination, efficiency, 505.
Image structure, 487.
Impedance, 30.
— matching, 355, 367, 476.
Indirectly heated valves, 85.
Induced E.M.F., 17.
Inductances in series or parallel, 20.
Induction, 15.
Infradyne, 300, 337.
Insulating materials, 141.
Interference, 338, 421.
Interlaced scanning, 519.
Intermediate frequency, 293, 300, 336.
Interrupted C.W., 168, 252.
Ionization, 73, 397.
Ionized layers, 401.
Ionosphere, 403, 413.
Iron cored coils, 337.

Jackson, Sir Henry, 394.

Karolous, 494.
Kennelly, 401.
Kerr cell, 517.
Key, morse, 67.
Keying, arc sets, 120.
— grid, 253.
— spark sets, 120.
— systems of, 252.
— telegraphic, 251.

Lamination, 41.
Langmuir, 71, 79.
Latour, 7.
Lee de Forest, Dr., 7, 74.
Lensed disk, 509.
Letcher wires, 443.
Light source, receiving, 505, 516.
Linear resonators, 146.
Lines of force, 15, 43.
Linkage, 18.
Load line, 364.
Loading of aerial, 51.
Locking of oscillators, 335.
Lodge, Sir Oliver, 4.
Logarithmic decrement, 40.
Long-lines oscillator, 145.
Loop aerial, 425.
Loose coupling, 61.
Losses, 41.
Loud speakers, 381.
— — coupling, 366.
Low-frequency transformers, 216.

McGaw, 150.
Magnetic coupling, 18, 59.
— deflection, 110.
— field, 15.
Magnetron, 149.
Marconi, 3, 9, 396.
— aerial, 47.
Marking wave, 120.
Master oscillator, 134.
May, 492.
Meissner, 7.
Metal rectifier, 191.
Mho, 30.
Microphone, carbon, 227.
— condenser, 231.
— double button, 229.
— moving coil, 231.
— piezo, 232.
— pressure, 351.
— velocity, 351.
Microphones, 230, 351.
Micro-wave oscillators, 149.
Micro-waves, 148.
Mihaly system, 511, 513.
Mirror drum, 511.
Modulated, C.W., 168, 252.
Modulation, cathode ray, 114, 528.
— low level, 254.
— telephony, 232, 323.
— television, 482.
Morrell, 149.
Mosaic system, 491.
Moving-coil speaker, 384.
Multiple valves, 98.
Multi-vibrator, 161.
Mutual conductance, 89.
— induction, 15, 19.

Natural frequency, 33, 39.
Navigation by radio, 432.
Negative cylinder, 104.
— resistance, 96, 118.
Neutralization, 282.
Night errors, 438.
Nipkow, Paul, 499.
Noctovision, 503.
Noise suppression, 315.

Octode, 96, 303.
Ohm's law, 15.

Open aerial, 51.
Oscillation, 38.
Oscillator, Hertzian, 43.
Output, L.F. power, 365.
Oxide coating, 79.

Padding condenser, 295.
Para-phase, 220.
Pedersen, 6.
Pentagrid, 303.
Pentode, 96, 368.
Periodic time, 47.
Permittivity, 23.
Phase, 26.
Photo-cells, 500.
Piezo-electricity, 152.
Plate, 71.
Polar curves, 353, 424, 447.
Polarization of light, 517.
Potential, 14.
Potentiometer, 170, 306.
Poulsen arc, 6, 118.
Power factor, 34.
— amplification, 360.
— output, 365.
— supply from mains, 184.
Power-grid detector, 176.
Pre-amplifier, 354.
Pre-selection, 294.
Privacy devices, 247.
Propagation of waves, 393.
Push-pull amplification, 219.
— — detection, 182.
— — power stages, 371.

'Q' of circuits, 263.
Quartz crystals, 152, 157.
— control, 151.
— crystal filter, 343.
— cutting, 154.
— oscillator circuits, 156.
Quenching, of detector, 275.
— of spark, 66.
Quiescent push-pull, 372.

Radiation, 42, 45, 453.
— resistance, 40, 49, 446.
Reactance, 17.
Reaction (regeneration), 124, 182, 268.
Receiver construction, 288.
— simple, 166, 268.
— television, 542.
Reception, 165, 258.
Rectification, 165.
— of A.C. mains, 184.
Rectifying valves, 100.
Reflection of waves, 406.
Reflectors, 451, 457.
Regulation, 187.
Rejector circuits, 261, 264.
Relays, 68.
Resistance coupling, 201.
Resistances, 17, 19.
Resonance, 33, 325.
— curves, 61, 63, 327.
Richardson, O. W., 71.
Robinson, Dr. J., 342, 410, 430.
Rochelle, salt, 153.
Rosing, 523.
Rotary spark gap, 66.
Round, Captain, 7, 51, 78.

Saturation, 72.
Scanning, 493, 499.
Scene, composition of, 483.
Schott effect, 271.
Scophany system, 515.
Screen, C.R., 106.
Screen-grid valve, 92, 285.
Screening, 42.
Second-channel, 293.
— detector, 306.
Secondary circuit, 59.
— emission, 95, 102.
Secrecy systems, 247.
Selectivity, 65, 260, 320.
Selenium cell, 492.
Sensitivity, 292.
Series, condensers, 25.
— modulation, 239.
Short-wave amplification, 287.
— — oscillators, 145.
— — valves, 141.
Side-band theory, 243, 321.
— — single, 341.
Sine wave, 26.
Skin effect, 42.
Skip-distance, 400.
Slope, 89.
Smoothing of H.T., 184.
Soft valve, 73, 77.
Space-charge, 72.
— waves, 397.
Spacing wave, 121.
Spark transmitter, 56.
Specific inductive capacity, 23.
Splash, side-band, 338.
Spotlight system, 503.
Square-law condenser, 297.
— detector, 171.
Standing waves, 442, 467.
Stenode principle, 341.
Stereoscopy, 483.
Stormer echoes, 405.
Superheterodyne, 292.
Super-regeneration, 273.
Suppressor-grid, 97.
Surface waves, 397.
Susceptance, 30.
Synchronization, 494, 530, 540.

Telephone condenser, 171.
Telephones, 166.
Television, 482.
— broadcast system, 536.
— cables, 544.
— cameras, 532.
— C.R. reception, 523, 542.
— frequencies, 495.
— images, 486.
— transmission, 500, 533.
Thermionic current, 70.
Thompson effect, 271.
Thoriated filament, 78.
Thyratron, 100, 530.
Time-base oscillator, 417, 529.
— constant, lag of screens, 107.
Tourmaline, 153.
Transceiver, 281.
Transformer, coupling, 213.
— line, 355.
— output, 366.
Transmission, continuous wave, 117.
— lines, 468.
— spark, 56.
— valve, 132.
Trapezium distortion, 532.
Triode, 74.
— power stage, 362.
Tuned anode coupling, 208.
— grid coupling, 215.
Tuning, 57.
— indicators, 316.
— of spark sets, 60.
— sharpness of, 64.
T.P.T.G., 137.
Tweeter, 387.

Ultra-short wave oscillators, 145.
— — propagation, 412.
Ultraudion, 137.
Universal receiver, 190.

Valve oscillators, 124.
Valves, 70.
— Fleming, 71.
Variable-mu valves, 309.
Variometer, 21.
Vector diagrams, 29.
Velocity, of waves, 47.
Voltage-doubler circuit, 192.
Volume control, 306.

Watson Watt, Dr., 412.
Wattless current, 34.
Wave form, 225, 233.
— of atmospherics, 418.
— of television, 539.
Wavelength, 48, 58.
— motion, 46.
— trap, 261, 264.
Wavemeter, 265.
Wehnelt, 80.
— cylinder, 104.
Weiller drum, 512.
Windom aerial, 476.
Wipe-out, 335.

X-cut crystals, 155.

Y-cut crystals, 155.

Zeehen effect, 335.
Zeppelin aerial, 466
Zincite, 169.
Zworikin, 11.

PRINTED IN
GREAT BRITAIN
BY
JARROLD AND SONS, LTD.
EMPIRE PRESS
NORWICH